Sparse and Low-Rank Representation
Lecture I: Motivation and Theory

Yi Ma

MSRA and UIUC

John Wright

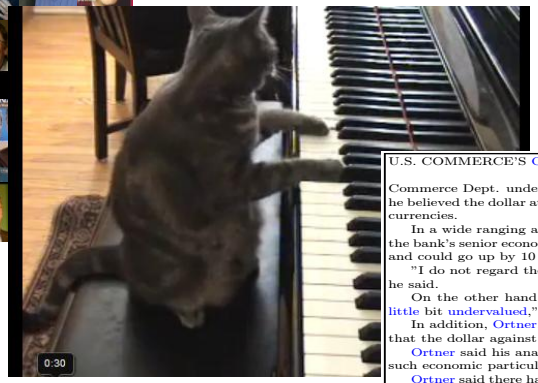
Columbia University

IPAM Computer Vision Summer School, August 5, 2013

CONTEXT – *Data increasingly massive, high-dimensional…*

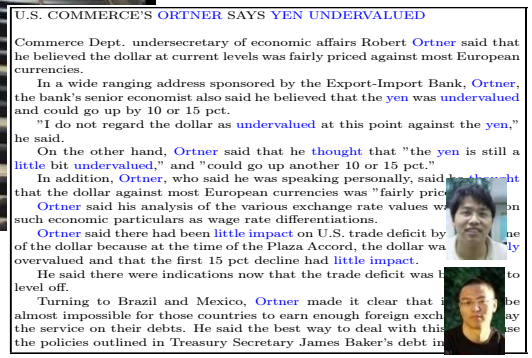
Images

- ***1M pixels***



Videos

- ***1B voxels***


U.S. COMMERCE'S ORTNER SAYS YEN UNDERVALUED

Commerce Dept. undersecretary of economic affairs Robert Ortner said that he believed the dollar at current levels was fairly priced against most European currencies.

In a wide ranging address sponsored by the Export-Import Bank, Ortner, the bank's senior economist also said he believed that the yen was undervalued and could go up by 10 or 15 pct.

"I do not regard the dollar as undervalued at this point against the yen," he said.

On the other hand, Ortner said that he thought that "the yen is still a little bit undervalued," and "could go up another 10 or 15 pct."

In addition, Ortner, who said he was speaking personally, said he thought that the dollar against most European currencies was "fairly priced."

Ortner said his analysis of the various exchange rate values was based on such economic particulars as wage rate differentiations.

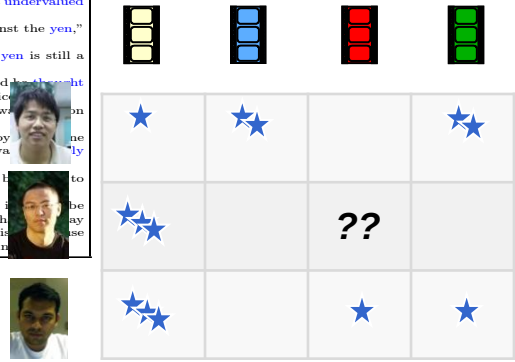
Ortner said there had been little impact on U.S. trade deficit by the decline of the dollar because at the time of the Plaza Accord, the dollar was greatly overvalued and that the first 15 pct decline had little impact.

He said there were indications now that the trade deficit was beginning to level off.

Turning to Brazil and Mexico, Ortner made it clear that it would be almost impossible for those countries to earn enough foreign exchange to pay the service on their debts. He said the best way to deal with this would be the policies outlined in Treasury Secretary James Baker's debt initiative.

Web data

- ***100B webpages***

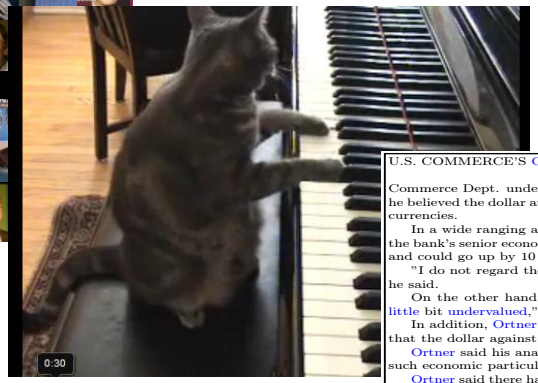


User data

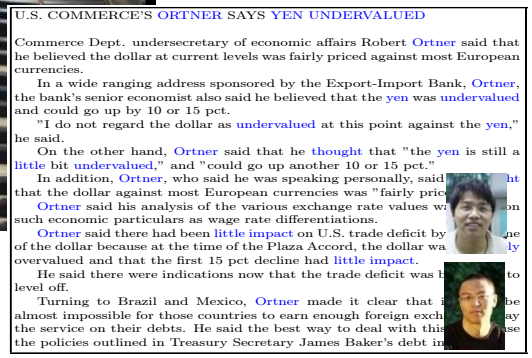
- ***1B users***

CONTEXT – *Discovering knowledge from data*

Images



Videos



U.S. COMMERCE'S ORTNER SAYS YEN UNDERVALUED

Commerce Dept. undersecretary of economic affairs Robert Ortner said that he believed the dollar at current levels was fairly priced against most European currencies.

In a wide ranging address sponsored by the Export-Import Bank, Ortner, the bank's senior economist also said he believed that the yen was undervalued and could go up by 10 or 15 pct.

"I do not regard the dollar as undervalued at this point against the yen," he said.

On the other hand, Ortner said that he thought that "the yen is still a little bit undervalued," and "could go up another 10 or 15 pct."

In addition, Ortner, who said he was speaking personally, said

Ortner said his analysis of the various exchange rate values w

such economic particulars as wage rate differentiations.

Ortner said there had been little impact on U.S. trade deficit by

of the dollar because at the time of the Plaza Accord, the dollar wa

overvalued and that the first 15 pct decline had little impact.

He said there were indications now that the trade deficit was b

level off.

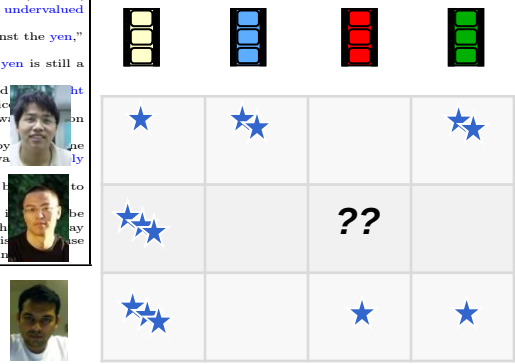
Turning to Brazil and Mexico, Ortner made it clear that

almost impossible for those countries to earn enough foreign exch

the service on their debts. He said the best way to deal with this

the policies outlined in Treasury Secretary James Baker's debt in

Web data



User data

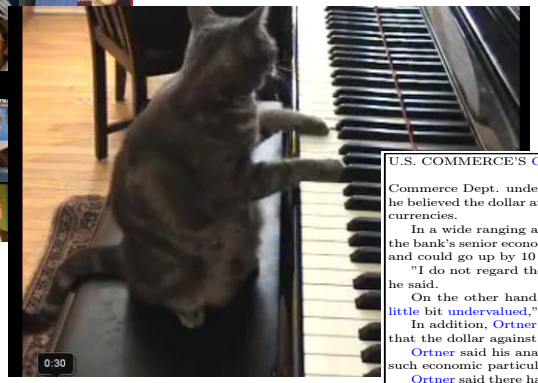
How to extract compact knowledge from such massive datasets?

CONTEXT – *Good solutions impact many applications*

Images

⇓

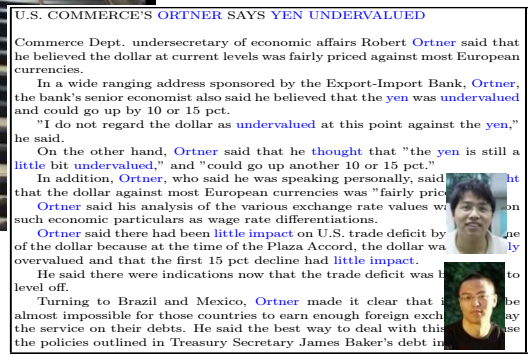
Compression
Denoising
Superresolution
Recognition…



Videos

⇓

Streaming
Tracking
Stabilization…



U.S. COMMERCE'S ORTNER SAYS YEN UNDERVALUED

Commerce Dept. undersecretary of economic affairs Robert Ortner said that he believed the dollar at current levels was fairly priced against most European currencies.

In a wide ranging address sponsored by the Export-Import Bank, Ortner, the bank's senior economist also said he believed that the yen was undervalued and could go up by 10 or 15 pct.

"I do not regard the dollar as undervalued at this point against the yen," he said.

On the other hand, Ortner said that he thought that "the yen is still a little bit undervalued," and "could go up another 10 or 15 pct."

In addition, Ortner, who said he was speaking personally, said

Ortner said his analysis of the various exchange rate values w

such economic particulars as wage rate differentiations.

Ortner said there had been little impact on U.S. trade deficit by

of the dollar because at the time of the Plaza Accord, the dollar wa

overvalued and that the first 15 pct decline had little impact.

He said there were indications now that the trade deficit was

level off.

Turning to Brazil and Mexico, Ortner made it clear that

almost impossible for those countries to earn enough foreign exch

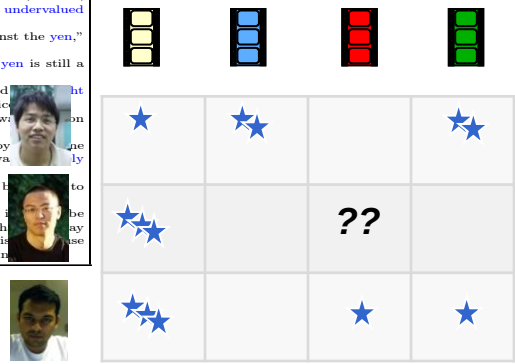
the service on their debts. He said the best way to deal with this

the policies outlined in Treasury Secretary James Baker's debt

Web data

⇓

Indexing
Ranking
Search…

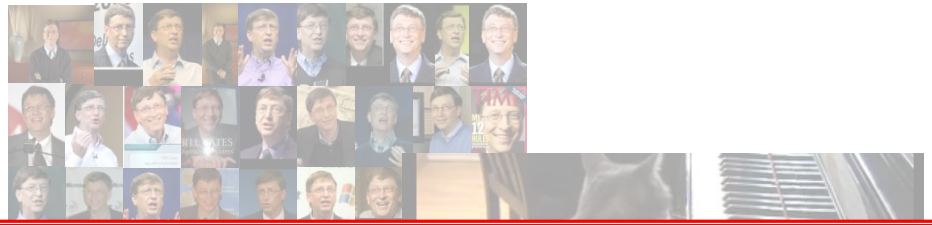


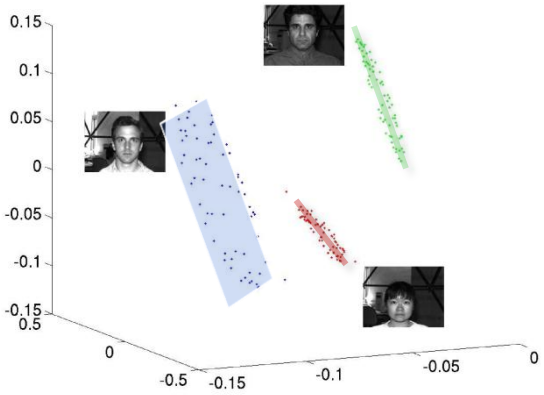
User data

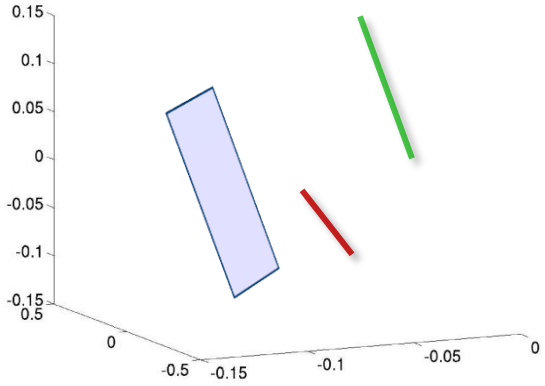
⇓

Clustering
Classification
Collaborative filtering…

Low-dimensional structures in high-dimensional data

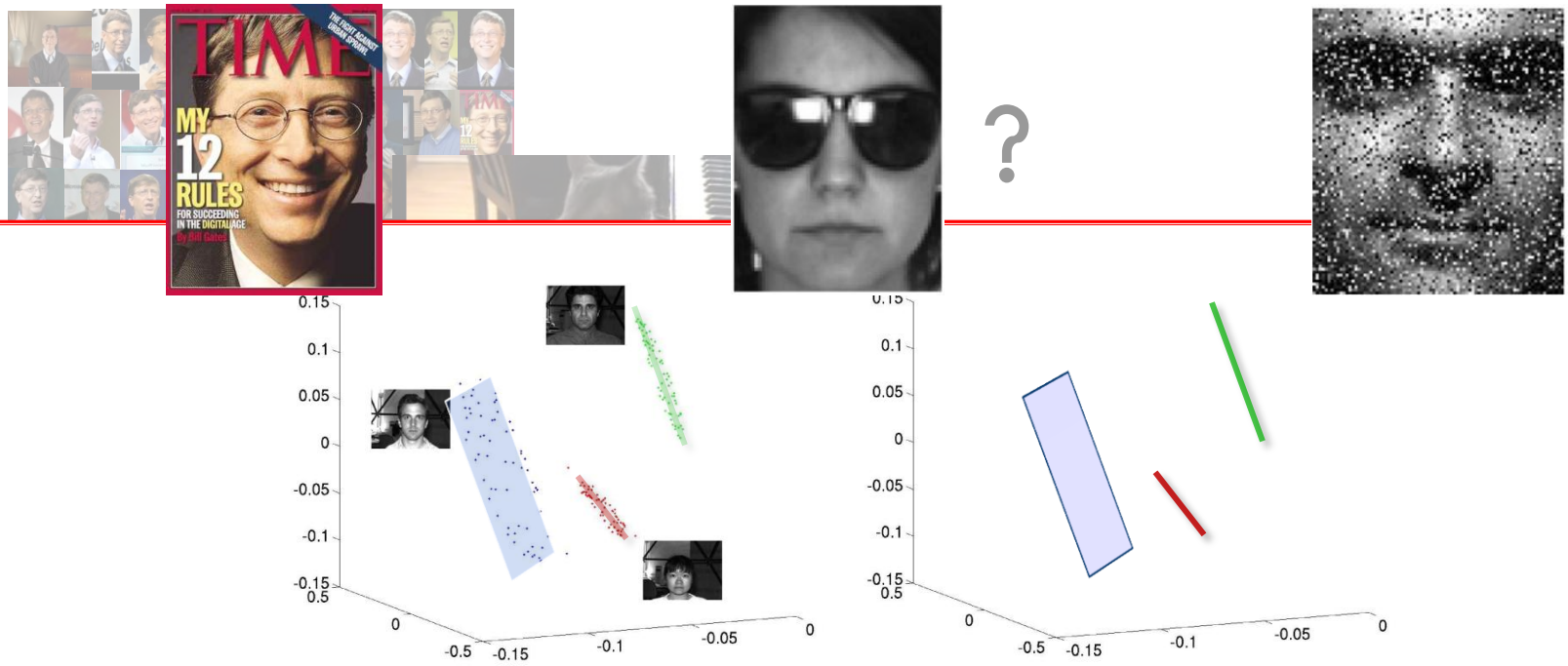






How can we learn and exploit low-dimensional structures in high-dimensional data?

But it is not so easy…



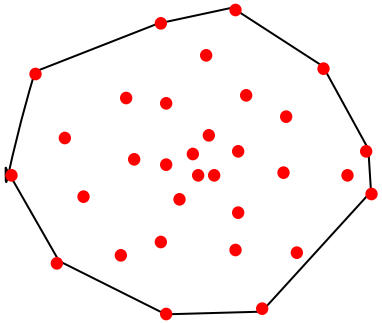
Real application data often contain missing observations, corruptions, or even malicious errors.

Classical methods (e.g., least squares, PCA) break down…

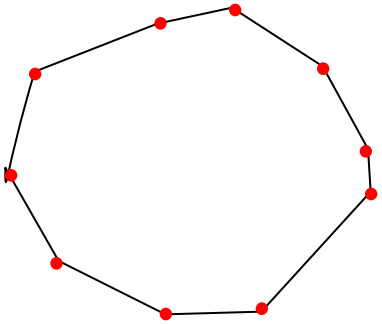
CONTEXT - *New Phenomena with High-Dimensional Data*

KEY CHALLENGE: efficiently and reliably recover low-dimensional structures from high-dimensional data, despite gross observation errors.

A sobering message: human intuition is severely limited in high-dimensional spaces:



Gaussian samples in 2D



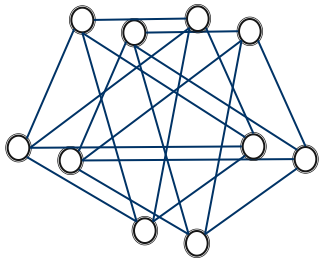
As dimension grows proportionally with the number of samples…

A **new regime** of geometry, statistics, and computation…

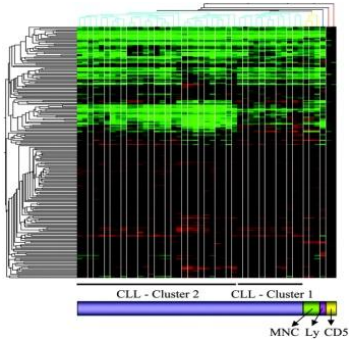
CONTEXT - *Massive High-Dimensional Data*

Recognition

Surveillance



Search and Ranking



Bioinformatics

The curse of dimensionality:
…increasingly demand inference with limited samples for very high-dimensional data.

The blessing of dimensionality:
… real data highly concentrate on low-dimensional, sparse, or degenerate structures in the high-dimensional space.

But **nothing is free**: *Gross errors and irrelevant measurements are now ubiquitous in massive cheap data.*

Everything old …

A long and rich history of robust estimation with error correction and missing data imputation:

R. J. Boscovich. ***De calculo probailitatum que respondent diversis valoribus summe errorum post plures observationes … , before 1756***

A. Legendre. ***Nouvelles methodes pour la determination des orbites des cometes*****, 1806**

C. Gauss. ***Theory of motion of heavenly bodies, 1809***

A. Beurling. ***Sur les integrales de Fourier absolument convergentes et leur application a une transformation functionelle*****, 1938**

B. Logan. ***Properties of High-Pass Signals*****, 1965**

$$L \quad x + \quad n$$

over-determined
+ dense, Gaussian

$$L \quad x + \quad e$$

underdetermined
+ sparse, Laplacian

⋮

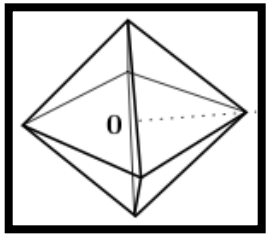
… IS NEW AGAIN

Today, robust estimation in high dimensions is more urgent, more tractable, and increasingly sharply understood.

Theory – high-dimensional geometry & statistics, measure concentration, combinatorics, coding theory…

Algorithms – large scale convex optimization, parallel and distributed computing….

Applications – massive data driven methods, sensing and hashing, denoising, superresolution, MRI, bioinformatics, image classification, recognition …



$$\min \|\boldsymbol{x}\|_\diamond \text{ s.t. } \boldsymbol{y} = A\boldsymbol{x}$$

The plan

Lecture I: Motivation and Theory

Lecture II: Data Modeling and Applications

Lecture III: Efficient Optimization for Low-Dimensional Models

Tomorrow: A Bayesian Perspective on Sparse Approximation

… + Q&A and discussion

Lecture I

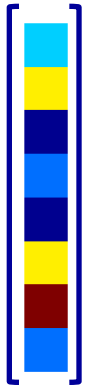
Theory of Sparse and Low-Rank Recovery

John Wright

Electrical Engineering

Columbia University

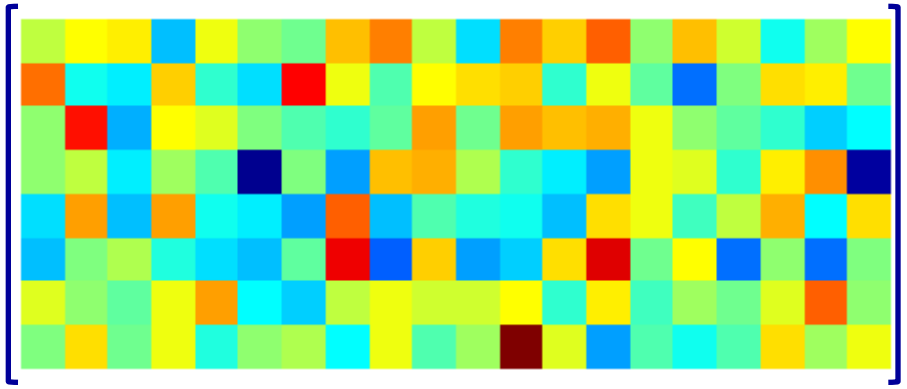
UNDERDETERMINED LINEAR SYSTEMS



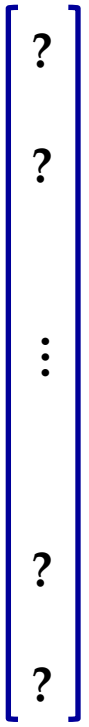
Observation

$\boldsymbol{y} \in \mathbb{R}^m$

$=$



$\boldsymbol{A} \in \mathbb{R}^{m \times n}$



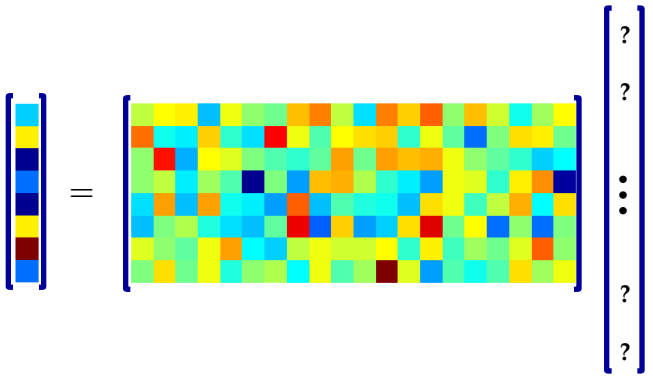
Unknown

$\boldsymbol{x} \in \mathbb{R}^n$

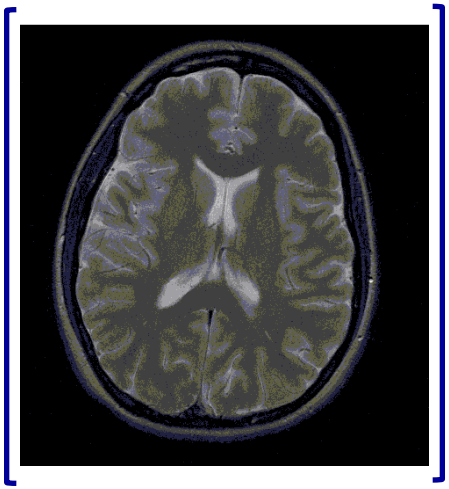
UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

$$y = Ax$$



Signal acquisition



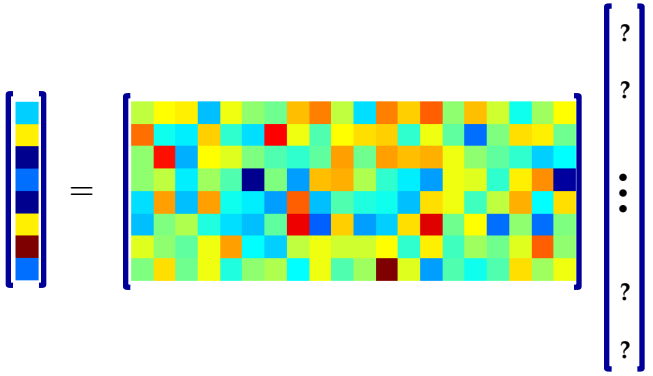
z

Image to be sensed

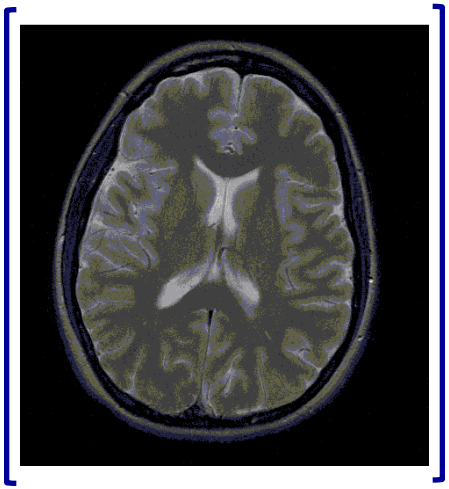
UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

$$y = Ax$$



Signal acquisition



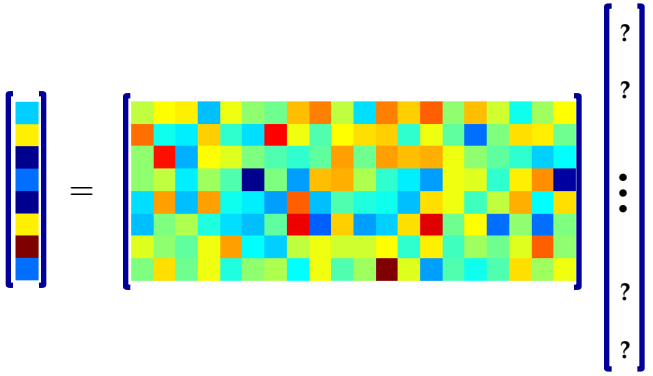
z

Image to be sensed

UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

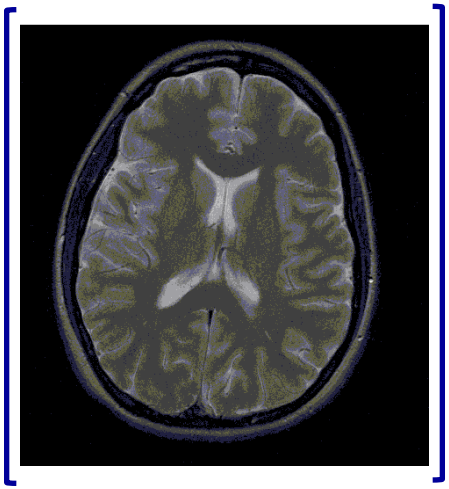
$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$



Signal acquisition

$$\boldsymbol{y}$$

$=$



$$\boldsymbol{z}$$

$$y_i = \int_{\boldsymbol{u}} \boldsymbol{z}(u) \exp(-2\pi j \boldsymbol{k}(t_i)^* \boldsymbol{u}) d\boldsymbol{u}$$

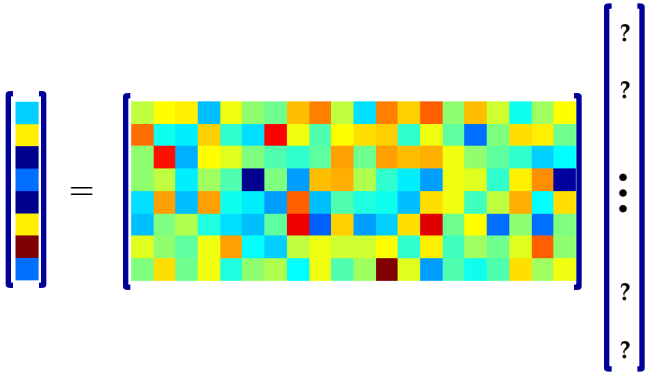
Observations are Fourier coefficients!

Image to be sensed

UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

$$y = Ax$$

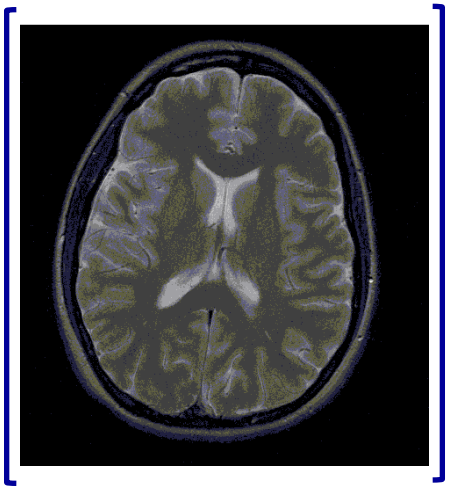


Signal acquisition

y  $=$ 

F_Ω

A few Fourier coefficients



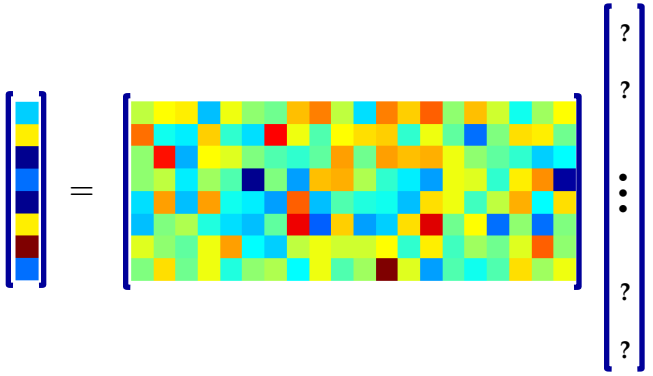
z

Image to be sensed

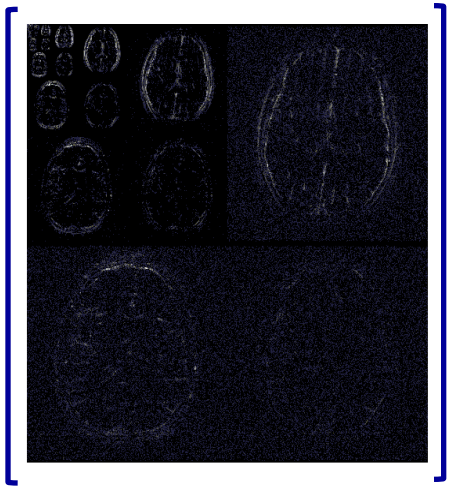
UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

$$y = Ax$$



Signal acquisition

y  $=$  Ψ 

F_Ω

x

A few Fourier coefficients

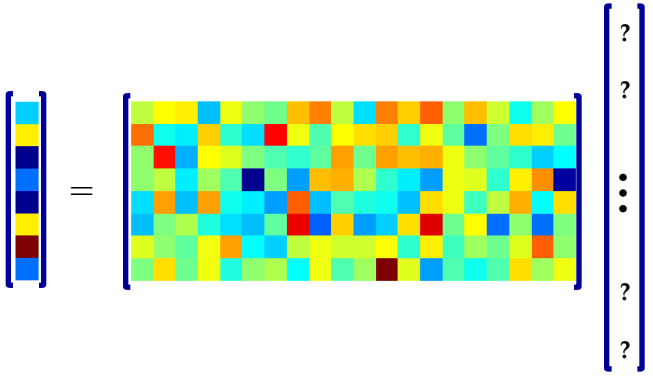
Wavelet coefficients: $z = \Psi x$

[Lustig, Donoho + Pauly '10] … brain image – Lustig '12

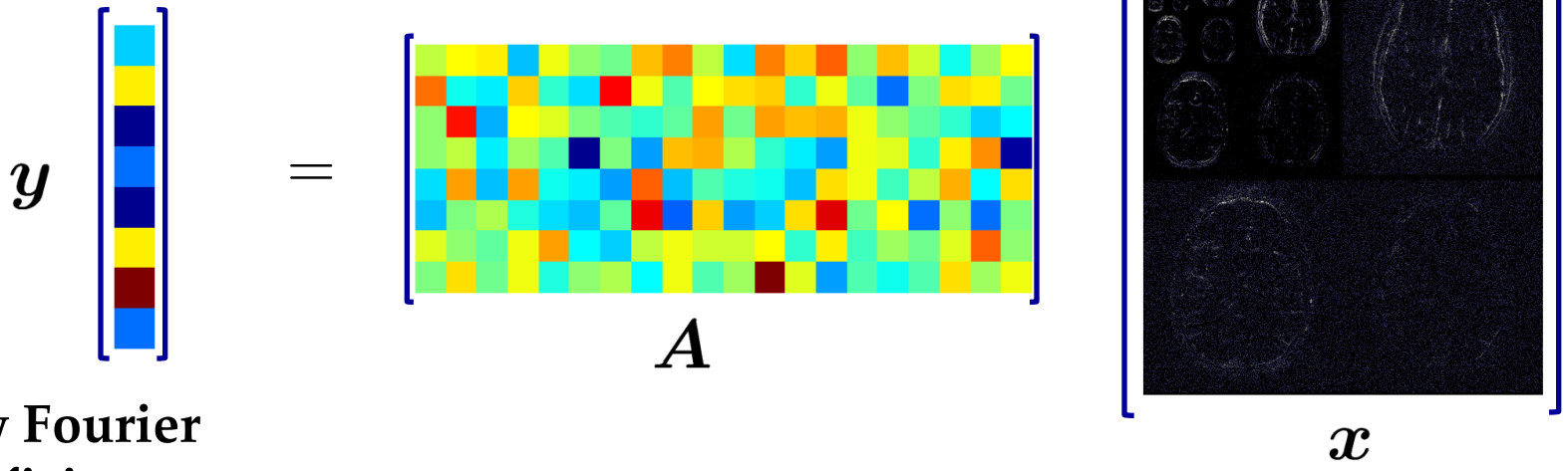
UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

$$y = Ax$$



Signal acquisition



A few Fourier coefficients

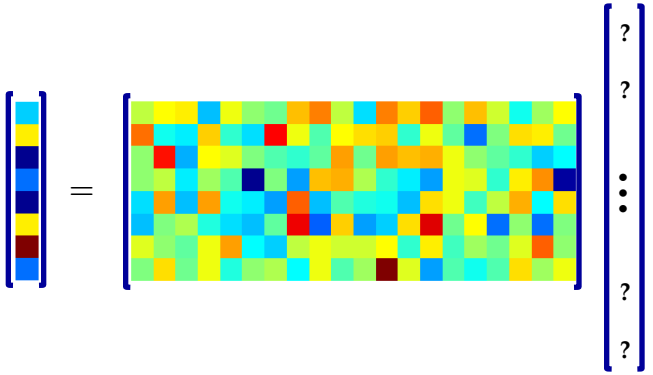
Wavelet coefficients

[Lustig, Donoho + Pauly '10] … brain image – Lustig '12

UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$



Compression

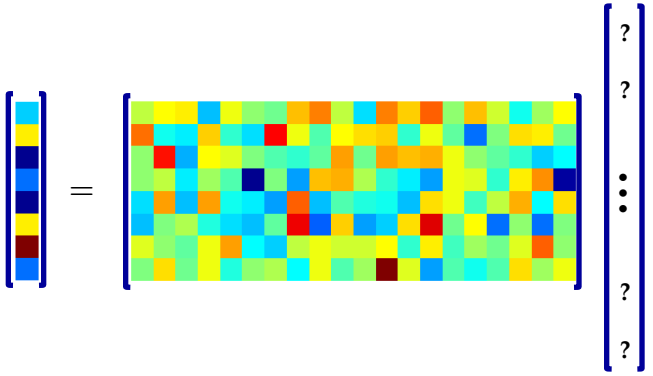
$\boldsymbol{y}$ 

Image to be compressed

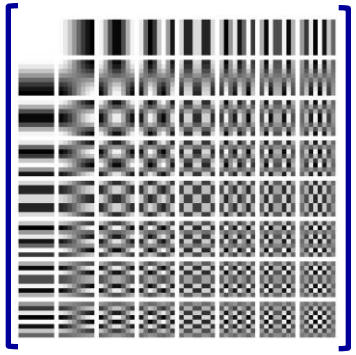
UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$



Compression – JPEG

$\boldsymbol{y}$  $\approx$  

(Patches of) … input image

$\boldsymbol{A}$ DCT basis

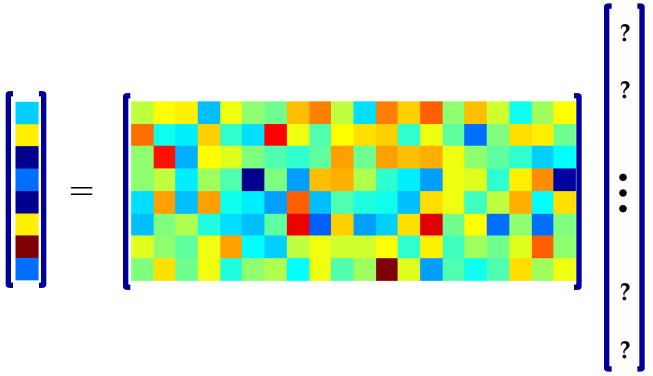
$\boldsymbol{x}$ coefficients

[Wallace '91]

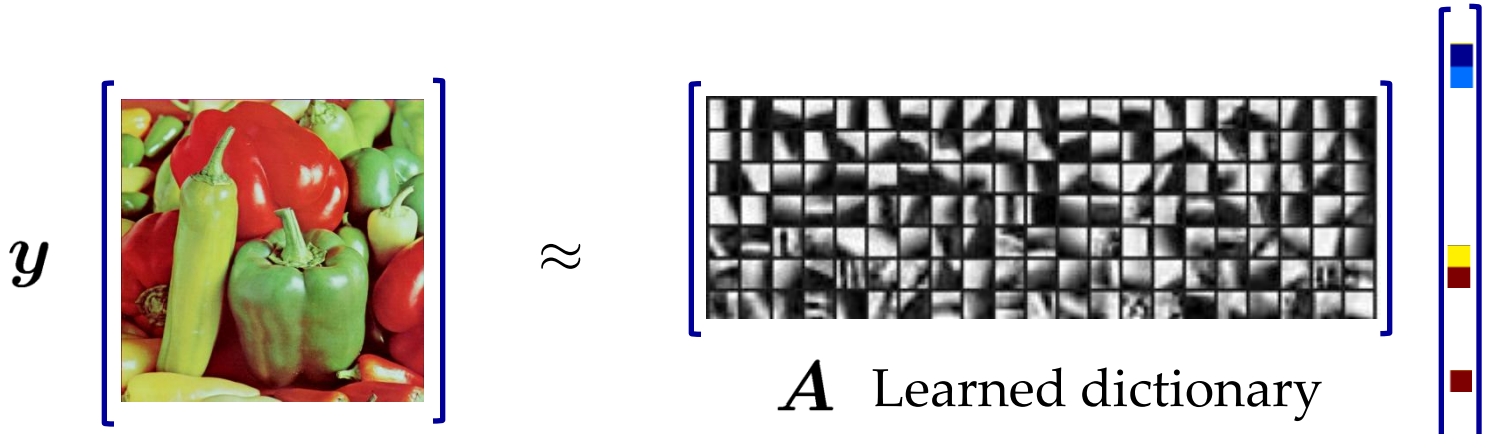
UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$



Compression – Learned dictionary



(Patches of) …
input image

$\boldsymbol{A}$ Learned dictionary

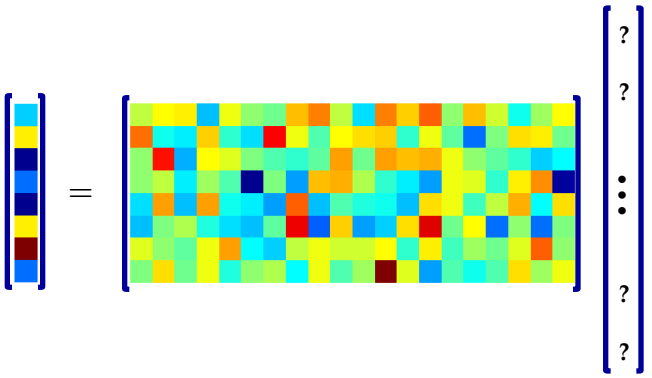
$\boldsymbol{x}$ coefficients

See [Elad+Bryt '08], [Horev et. Al., '12] … Image: [Aharon+Elad '05]

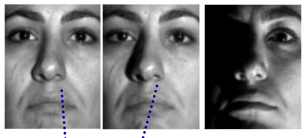
UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

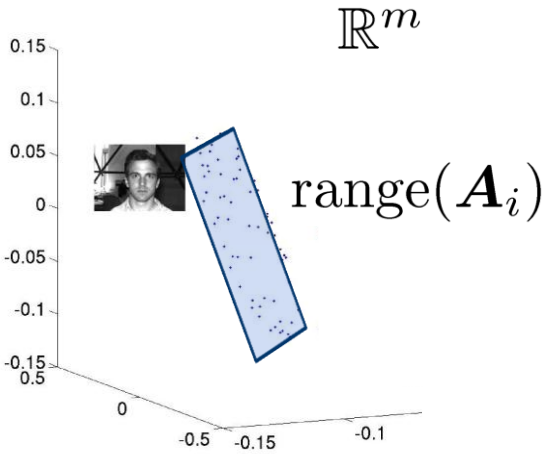
$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$



Recognition



$$\boldsymbol{A}_i = [\ |\ |\ \ldots\ ] \in \mathbb{R}^{m \times n_i}$$



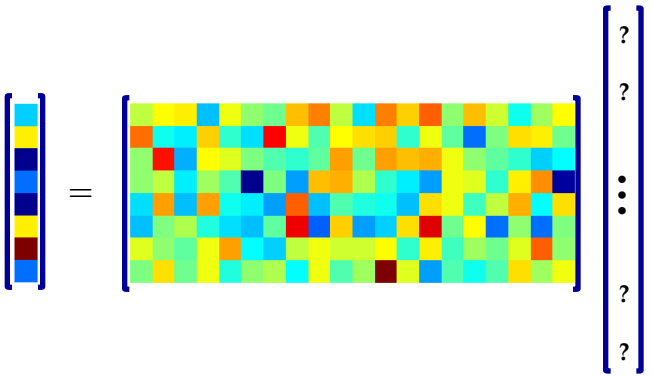
Linear subspace model for images of **same face** under **varying lighting.**

[Basri+Jacobs '03], [Ramamoorthi '03], [Belhumeur+Kriegman '96]

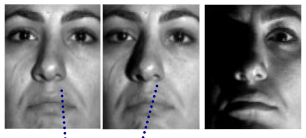
UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

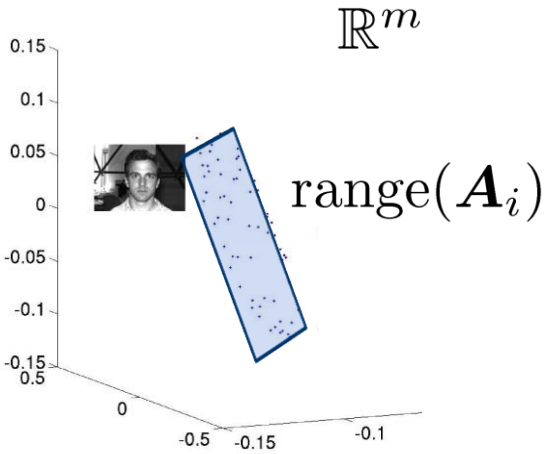
$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$



Recognition



$$\boldsymbol{A}_i = [\ \ |\ \ |\ \ \ldots\ ] \in \mathbb{R}^{m \times n_i}$$

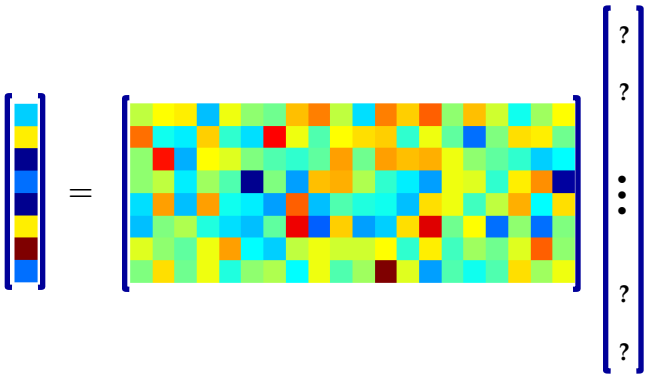


$$\boldsymbol{y} \approx x_{i,1} + x_{i,2} + \ldots + x_{i,n} = \boldsymbol{A}_i \boldsymbol{x}_i$$

UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

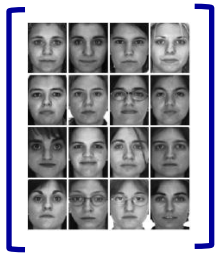


Recognition

$\boldsymbol{y} \in \mathbb{R}^m$

Test image

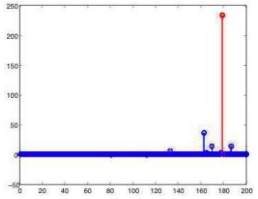
$=$



$\boldsymbol{A} = [\boldsymbol{A}_1 \mid \boldsymbol{A}_2 \mid \cdots \mid \boldsymbol{A}_k]$

Combined training dictionary

$\times$



$\boldsymbol{x} \in \mathbb{R}^n$

coefficients

$+$

$\boldsymbol{e} \in \mathbb{R}^m$

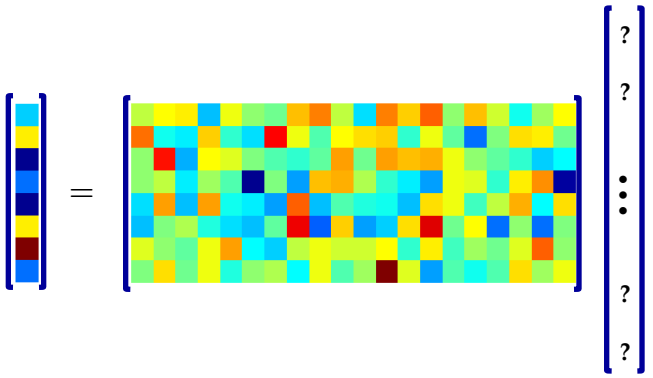
corruption, occlusion

[W., Yang, Ganesh, Sastry, Ma '09]

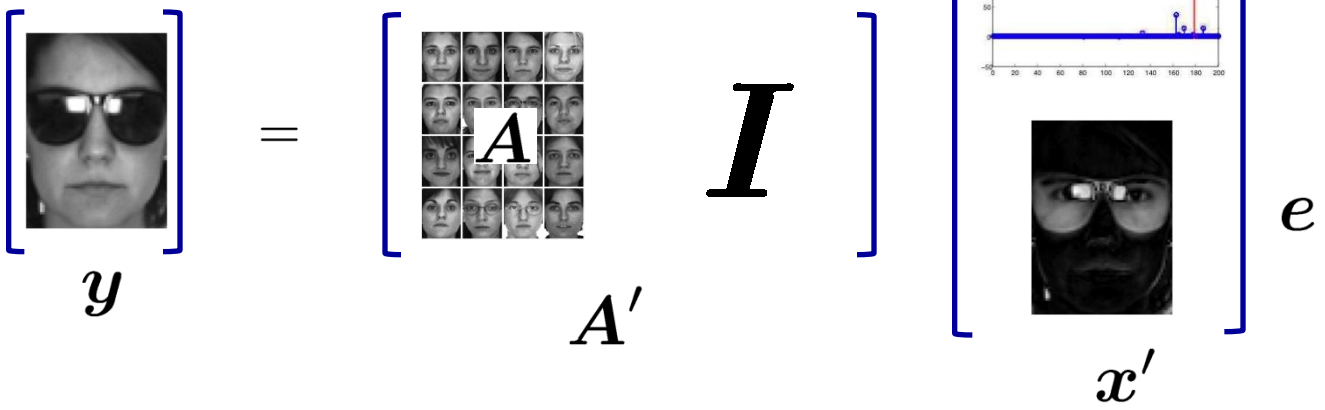
UNDERDETERMINED LINEAR SYSTEMS

Underdetermined system

$$y = Ax$$

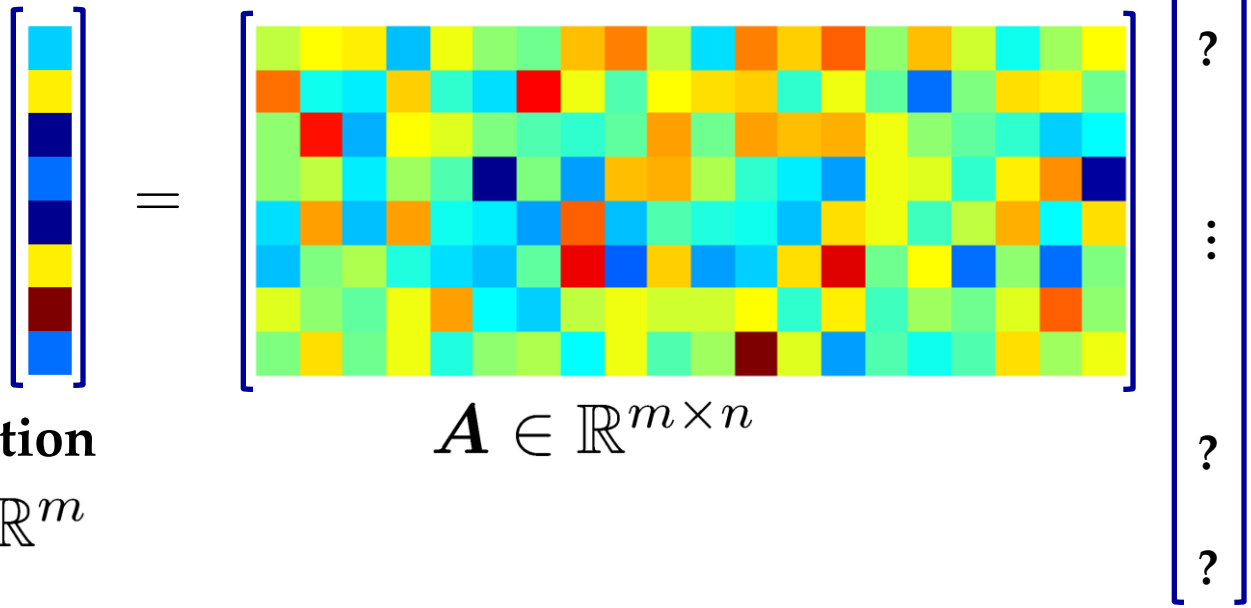


Recognition



One large underdetermined system: $y = A'x'$

UNDERDETERMINED LINEAR SYSTEMS



Observation

$\boldsymbol{y} \in \mathbb{R}^m$

$\boldsymbol{A} \in \mathbb{R}^{m \times n}$

Unknown

$\boldsymbol{x} \in \mathbb{R}^n$

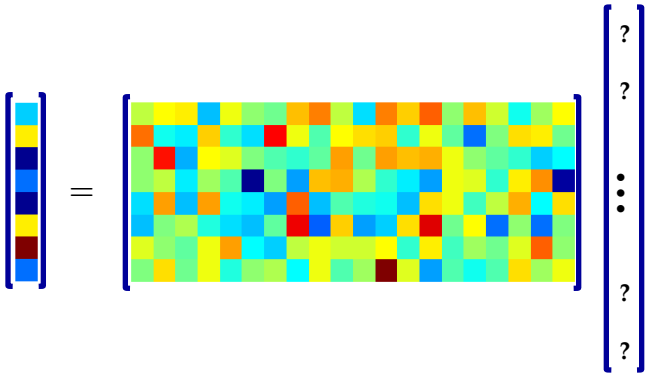
In all of these examples, $\underbrace{m}_{\texttt{\#observations}} \ll \underbrace{n}_{\texttt{\#unknowns}}$.

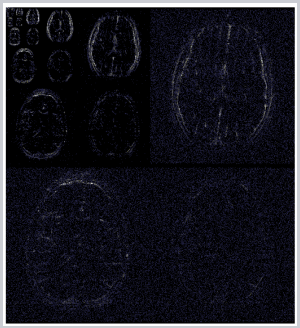
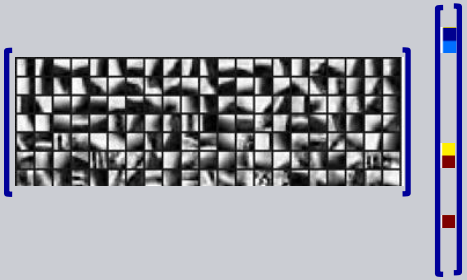
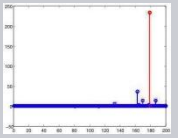
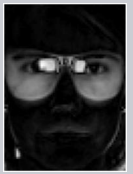
Solution is **not unique** … is there any hope?

WHAT DO WE KNOW ABOUT x ?

Underdetermined system

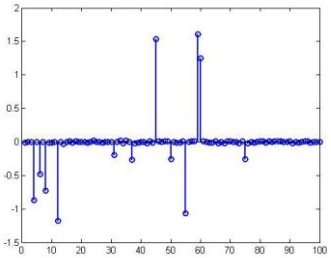
$$y = Ax$$



Signal acquisition	Image compression	Face Recognition
		 
$x^\star$ contains **just a few** significant wavelet coefficients.	$x^\star$ uses **just a few** dictionary elements.	$x^\star$ uses **just a few** training faces.  $e^\star$ corrects **a few** gross errors.

SPARSITY – More formally

A vector $\boldsymbol{x} \in \mathbb{R}^n$ is **sparse** if only a few entries are nonzero:

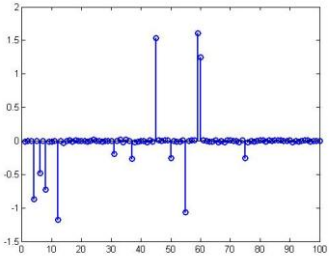


The **number of nonzeros** is called the ℓ^0**-"norm"** of $\boldsymbol{x}$:

$$\|\boldsymbol{x}\|_0 \doteq \#\{i \mid x_i \neq 0\}.$$

SPARSITY – More formally

A vector $\boldsymbol{x} \in \mathbb{R}^n$ is **sparse** if only a few entries are nonzero:



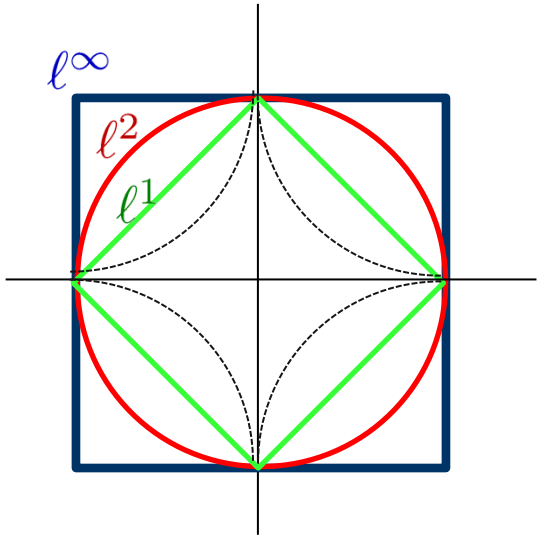
The **number of nonzeros** is called the ℓ^0-**"norm"** of $\boldsymbol{x}$:

$$\|\boldsymbol{x}\|_0 \doteq \#\{i \mid x_i \neq 0\}.$$

Geometrically

$$\|\boldsymbol{x}\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$$

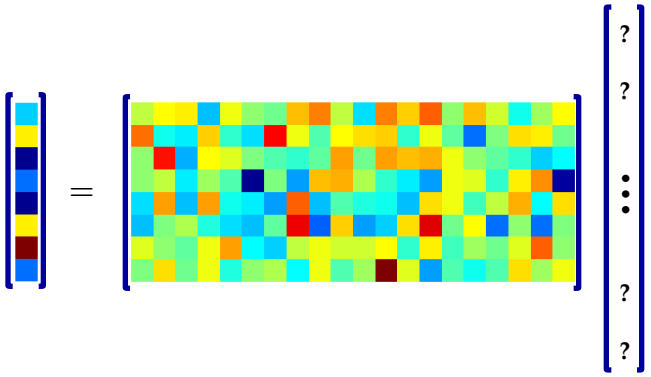
$$\|\boldsymbol{x}\|_0 = \lim_{p \searrow 0} \|\boldsymbol{x}\|_p^p.$$



THE SPARSEST SOLUTION

Underdetermined system

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$



Look for the sparsest $\boldsymbol{x}$ that agrees with our observation:

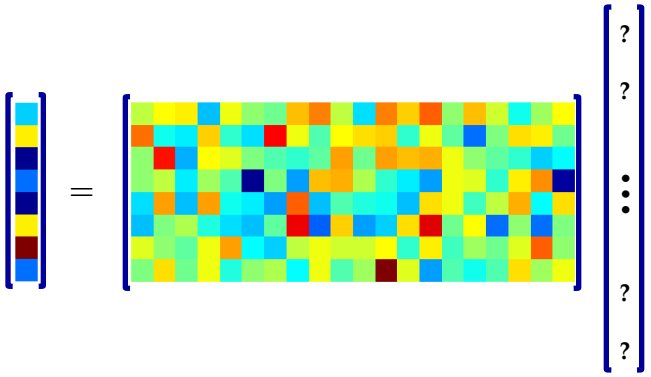
$$\text{minimize} \quad \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

[Demo]

THE SPARSEST SOLUTION

Underdetermined system

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$



Look for the sparsest $\boldsymbol{x}$ that agrees with our observation:

$$\text{minimize} \quad \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

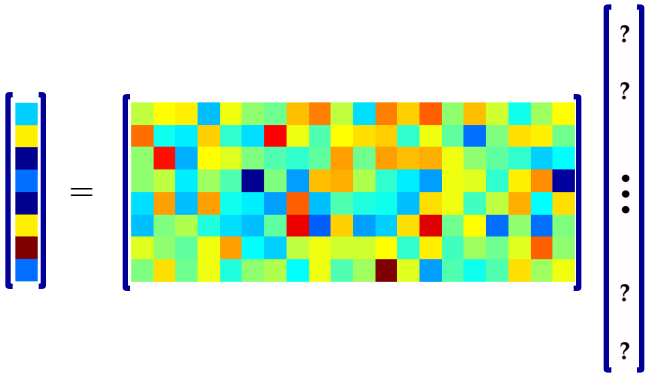
Theorem 1 (Gorodnitsky+Rao '97) .
Suppose $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$, *and let* $k = \|\boldsymbol{x}_0\|_0$. *If* $\text{null}(\boldsymbol{A})$ *contains no* $2k$*-sparse vectors,* $\boldsymbol{x}_0$ *is the unique optimal solution to*

$$\text{minimize } \|\boldsymbol{x}\|_0 \quad \textit{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

THE SPARSEST SOLUTION

Underdetermined system

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$



Look for the sparsest $\boldsymbol{x}$ that agrees with our observation:

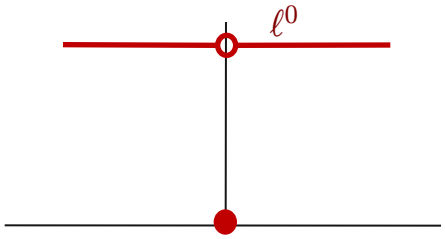
$$\text{minimize} \quad \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

INTRACTABLE

RELAX!

$$\text{minimize} \quad \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

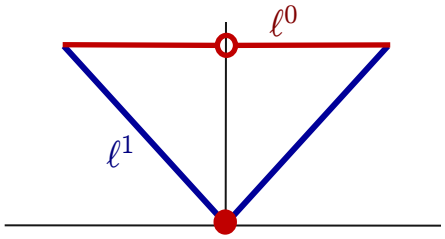
The cardinality $\|\boldsymbol{x}\|_0$ is **nonconvex**:



RELAX!

$$\text{minimize} \quad \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

The cardinality $\|\boldsymbol{x}\|_0$ is **nonconvex**:



Its **convex envelope*** is

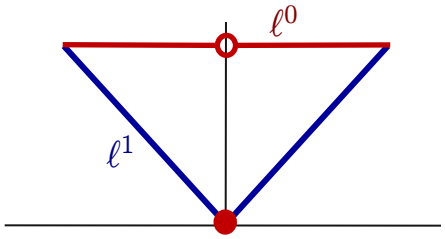
the ℓ^1**norm:** $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$

* Over the set $\{\boldsymbol{x} \mid |x_i| \leq 1 \; \forall \; i\}$

RELAX!

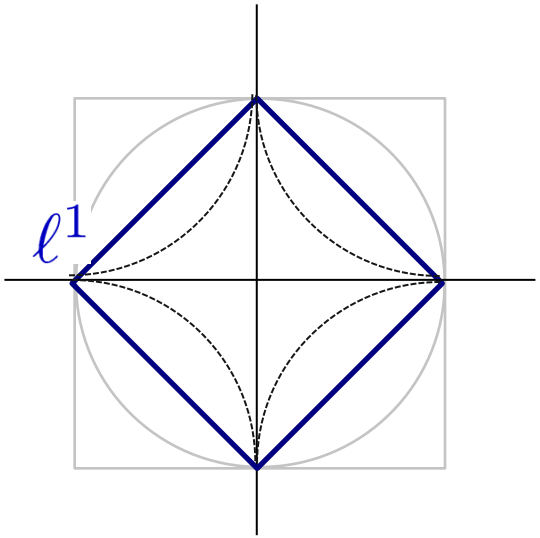
$$\text{minimize} \quad \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

The cardinality $\|\boldsymbol{x}\|_0$ is **nonconvex**:



Its **convex envelope*** is

the ℓ^1**norm:** $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$



* Over the set $\{\boldsymbol{x} \mid |x_i| \leq 1 \; \forall \; i\}$

RELAX!

$$\text{minimize} \quad \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

NP-hard, hard to appx.
[Natarjan '95],
[Amaldi+Kann '97]

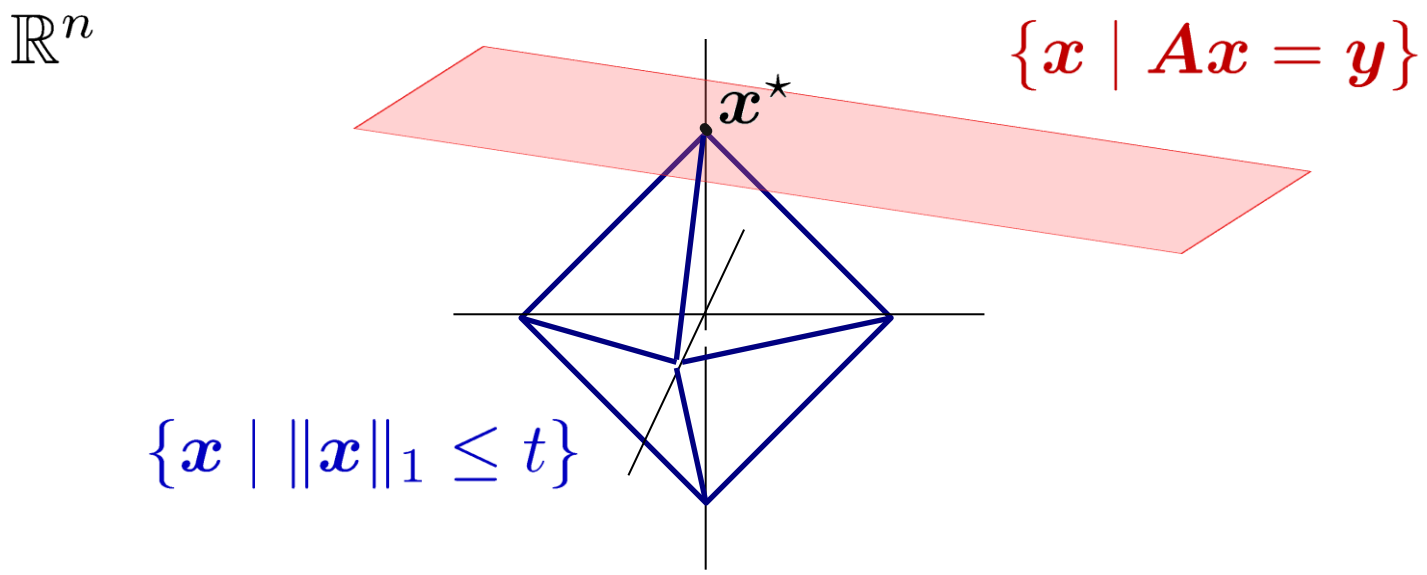
$$\text{minimize} \quad \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

Efficiently solvable
– Lecture 3!

Have we lost anything? [demo]

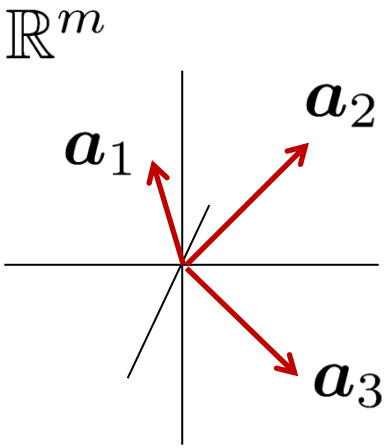
WHY DOES THIS WORK? Geometric intuition

$$\text{minimize} \quad \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$



WHY DOES THIS WORK? More formally…

We see: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \sum_{i \in \mathrm{supp}(\boldsymbol{x})} \boldsymbol{a}_i x_i$



Intuition: Recovering $\boldsymbol{x}$ is "easier" if the $\boldsymbol{a}_i$ are not too similar…

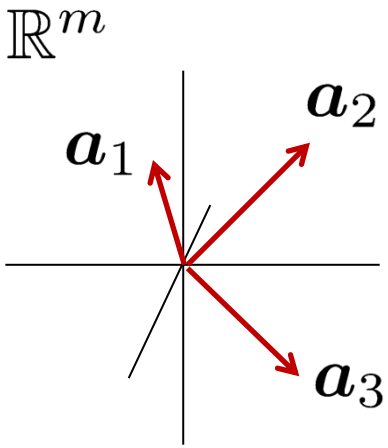
Mutual coherence $\mu(\boldsymbol{A}) \doteq \max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|$

Smaller is better!

WHY DOES THIS WORK? More formally…

Mutual coherence

$$\mu(\boldsymbol{A}) \doteq \max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|$$



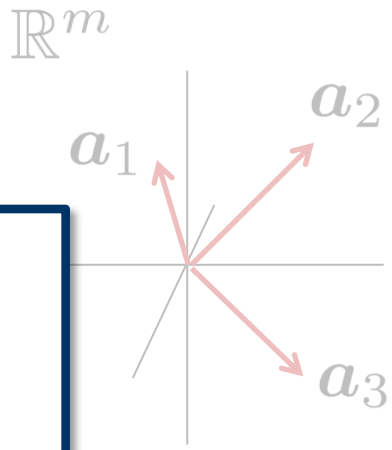
Theorem 2 (Gribonval+Nielsen '03, Donoho+Elad '03) .
Suppose $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ *with*

$$\|\boldsymbol{x}_0\|_0 < \tfrac{1}{2}(1 + 1/\mu(\boldsymbol{A})).$$

Then $\boldsymbol{x}_0$ *is the unique optimal solution to*

$$\text{minimize } \|\boldsymbol{x}\|_1 \quad \textit{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

WHY DOES THIS WORK? More formally…



Mutual coherence

The target solution $\boldsymbol{x}_0$ is **sufficiently structured** (sparse!).

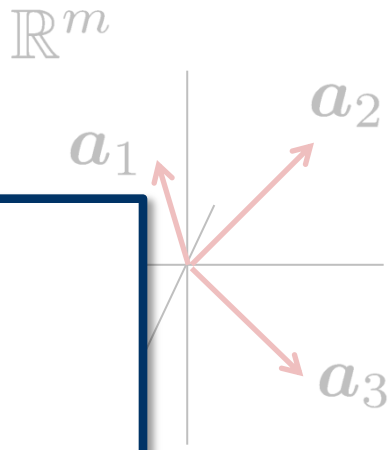
Theorem 2 (Gribonval+Nielsen '03, Donoho+Elad '03).
Suppose $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ *with*

$$\|\boldsymbol{x}_0\|_0 < \tfrac{1}{2}(1 + 1/\mu(\boldsymbol{A})).$$

Then $\boldsymbol{x}_0$ *is the unique optimal solution to*

$$\text{minimize } \|\boldsymbol{x}\|_1 \quad \textit{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

WHY DOES THIS WORK? More formally…



Mutual coherence

The matrix $\boldsymbol{A}$ is **incoherent** – and so, preserves sparse $\boldsymbol{x}$.

Theorem 2 (Gribonval+Nielsen '03, Donoho+Elad '03).
Suppose $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ *with*

$$\|\boldsymbol{x}_0\|_0 < \tfrac{1}{2}(1 + 1/\mu(\boldsymbol{A})).$$

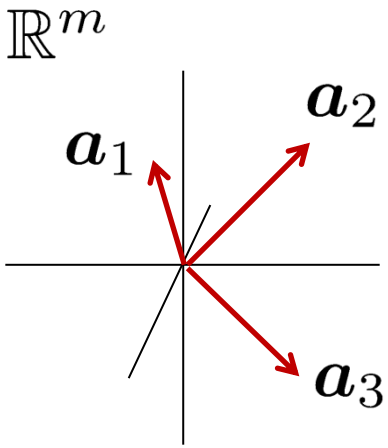
Then $\boldsymbol{x}_0$ *is the unique optimal solution to*

$$\text{minimize } \|\boldsymbol{x}\|_1 \quad \textit{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

WHY DOES THIS WORK? More formally…

Mutual coherence

$$\mu(\boldsymbol{A}) \doteq \max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|$$



Theorem 2 (Gribonval+Nielsen '03, Donoho+Elad '03) .
Suppose $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ *with*

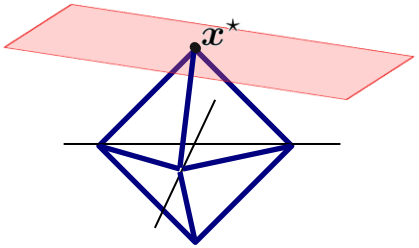
$$\|\boldsymbol{x}_0\|_0 < \tfrac{1}{2}(1 + 1/\mu(\boldsymbol{A})).$$

Then $\boldsymbol{x}_0$ *is the unique optimal solution to*

$$\text{minimize } \|\boldsymbol{x}\|_1 \quad \textit{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

WHY CARE ABOUT THE THEORY?

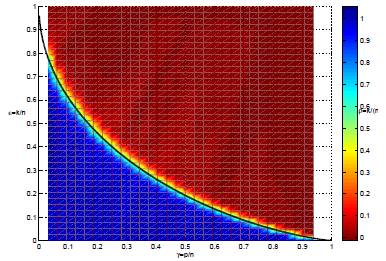
Motivates applications



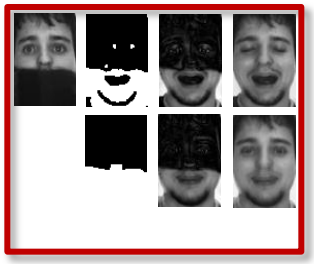
… but be careful: need to justify (and modify) the basic models [Lecture 2].

Template for stronger results

… predictions can be very sharp in high dimensions.

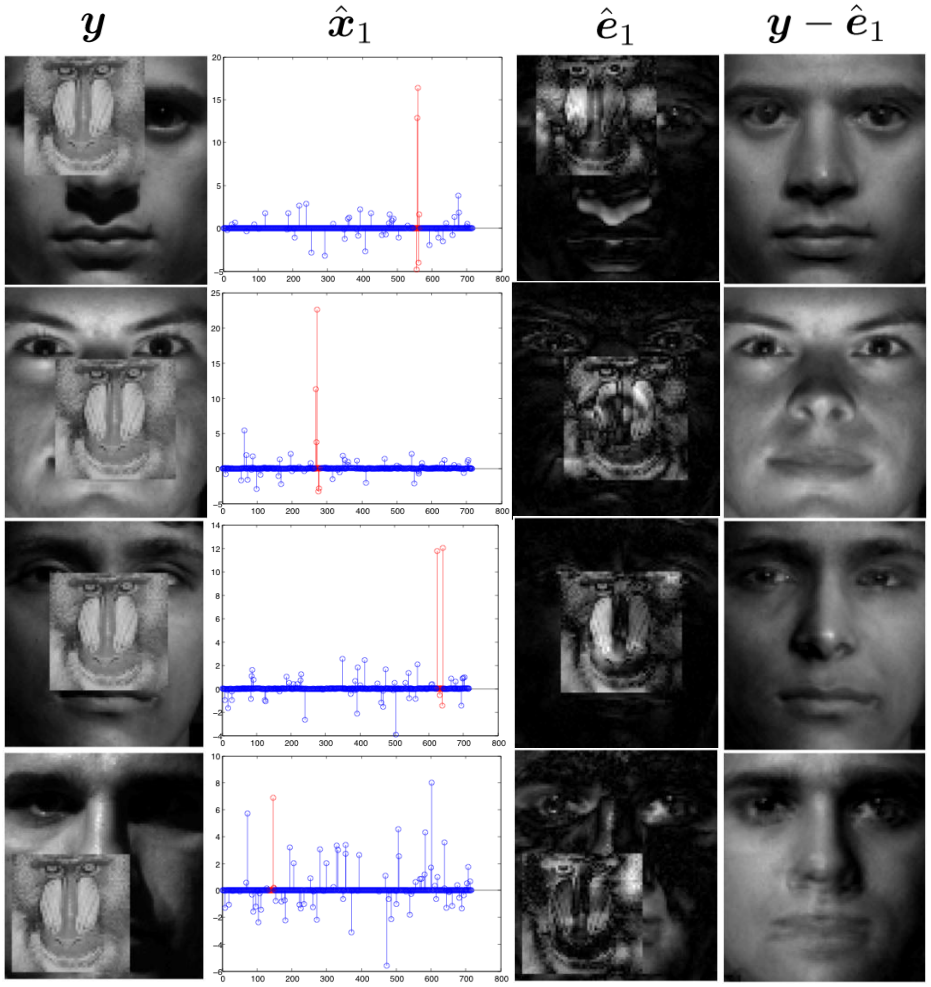


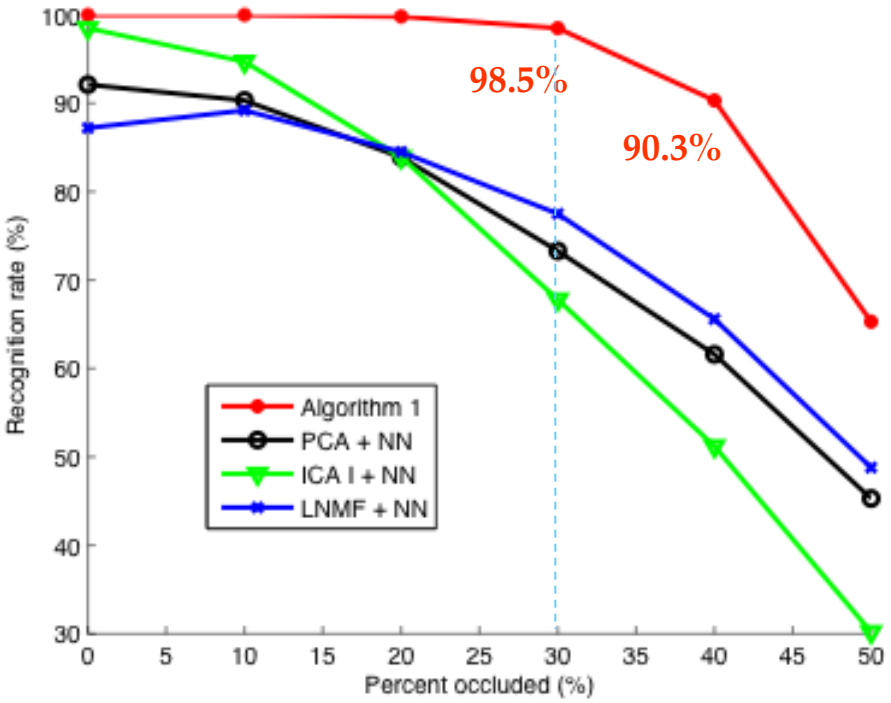
Generalizes to many other types of low-dimensional structure



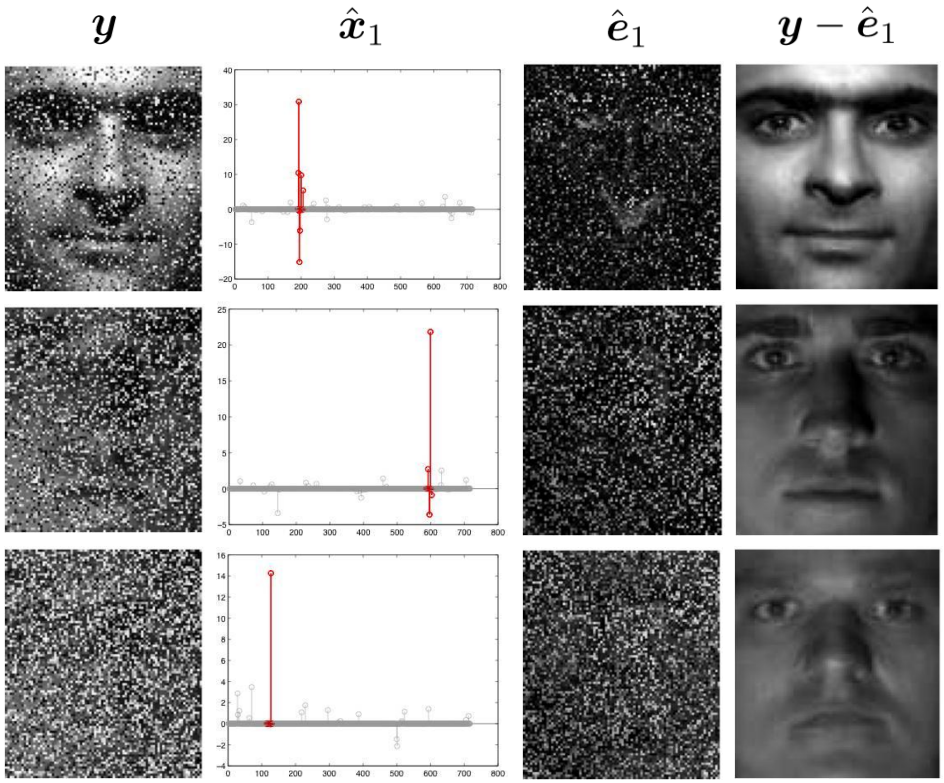
… structured sparsity, low-rank recovery [later, Lecture 2]

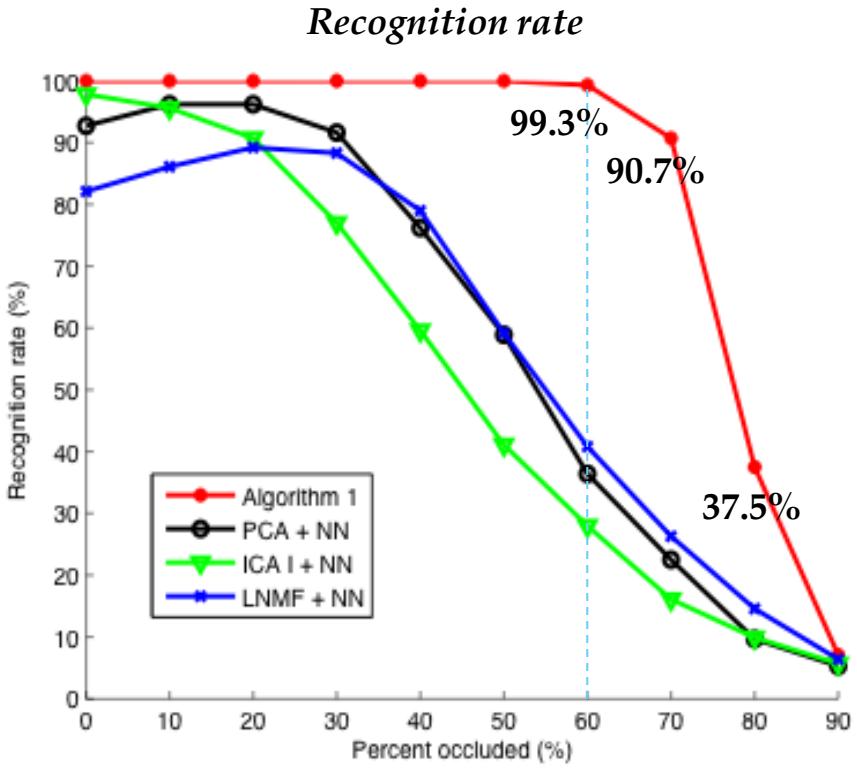
MOTIVATING APPLICATIONS – Face Recognition





MOTIVATING APPLICATIONS – Face Recognition

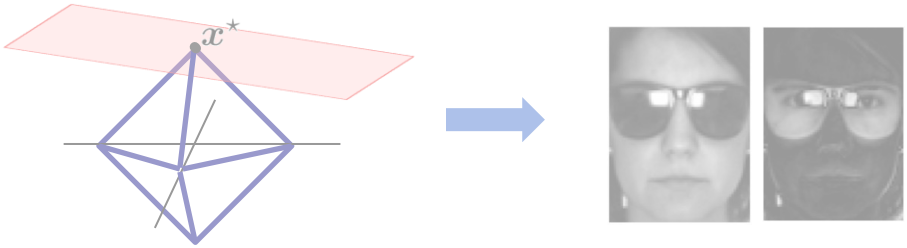




More practicalities in Lecture 2…

WHY CARE ABOUT THE THEORY?

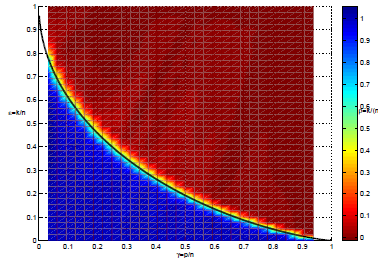
Motivates applications



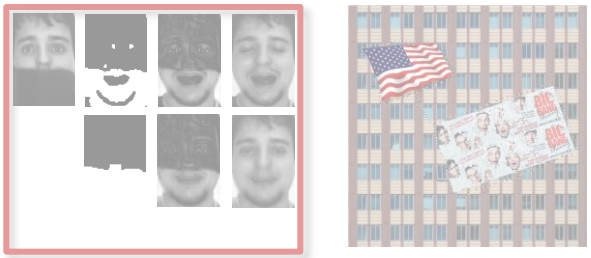
… but be careful: need to justify (and modify) the basic models [Lecture 2].

Template for stronger results

… predictions can be very sharp in high dimensions.



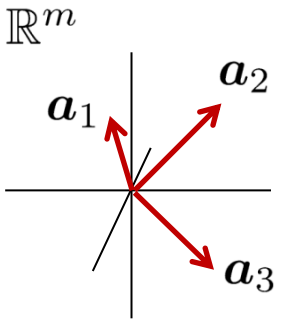
Generalizes to many other types of low-dimensional structure



… structured sparsity, low-rank recovery [later, Lecture 2]

LIMITATIONS OF COHERENCE?

For any $m \times n$ $\boldsymbol{A}$ $\quad \mu(\boldsymbol{A}) \geq \sqrt{\frac{n-m}{m(n-1)}}$

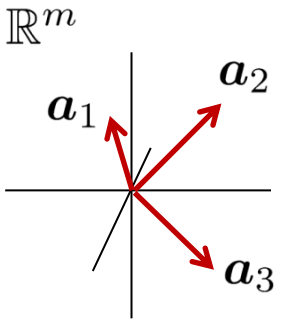


Prev. result therefore requires

$$\|\boldsymbol{x}_0\|_0 \; < \; \tfrac{1}{2}(1 + \mu(\boldsymbol{A})^{-1}) \; = \; O(\sqrt{m})$$

LIMITATIONS OF COHERENCE?

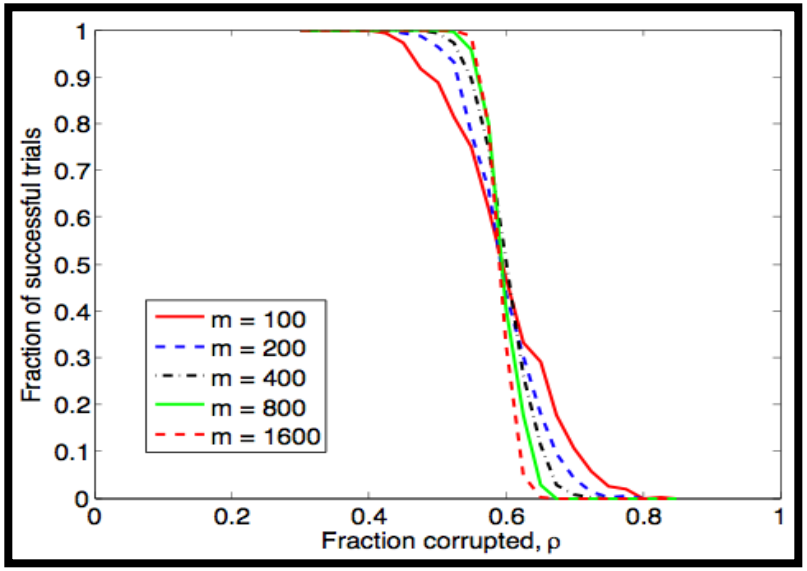
For any $m \times n$ $\boldsymbol{A}$ $\quad \mu(\boldsymbol{A}) \geq \sqrt{\frac{n-m}{m(n-1)}}$



Prev. result therefore requires

$$\|\boldsymbol{x}_0\|_0 \; < \; \tfrac{1}{2}(1 + \mu(\boldsymbol{A})^{-1}) \; = \; O(\sqrt{m})$$

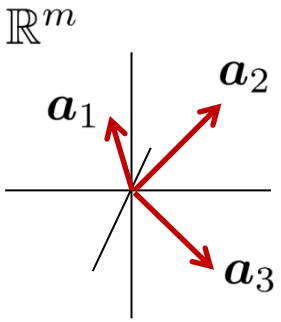
Truth is often **much better**:



Plot: Fraction of correct recovery vs. fraction of nonzeros $\|\boldsymbol{x}_0\|_0/m$

LIMITATIONS OF COHERENCE?

For any $m \times n$ $\boldsymbol{A}$ $\quad \mu(\boldsymbol{A}) \geq \sqrt{\frac{n-m}{m(n-1)}}$



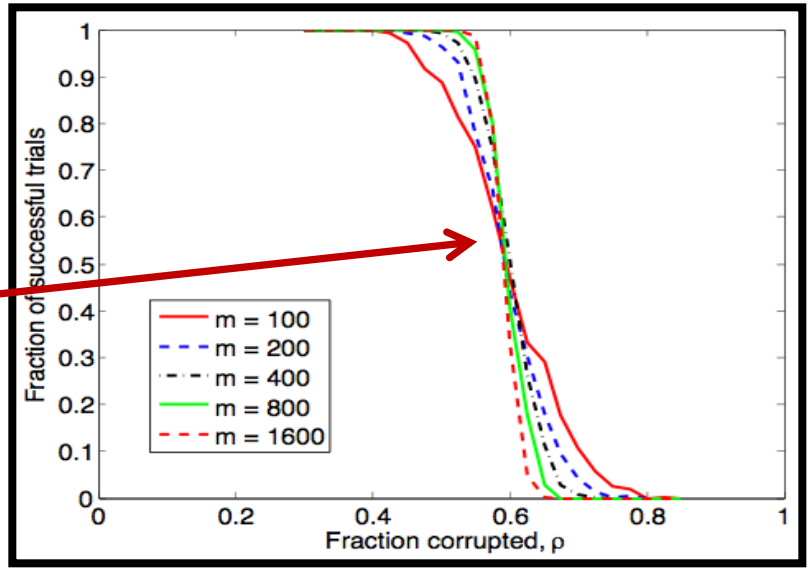
Prev. result therefore requires

$$\|\boldsymbol{x}_0\|_0 \; < \; \tfrac{1}{2}(1 + \mu(\boldsymbol{A})^{-1}) \; = \; O(\sqrt{m})$$

Truth is often **much better**:

Phase transition at

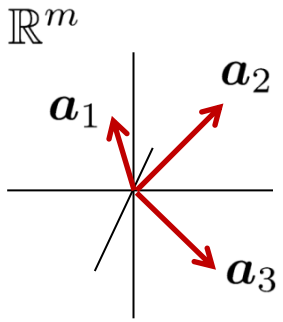
$$\|\boldsymbol{x}_0\|_0 = \alpha^\star m$$



Plot: Fraction of correct recovery vs. fraction of nonzeros $\|\boldsymbol{x}_0\|_0/m$

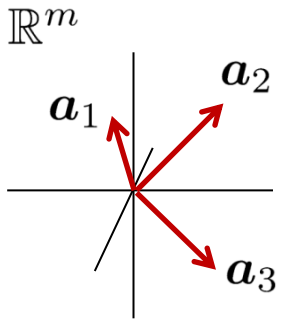
STRENGTHENING THE BOUND – the RIP

Incoherence: **Each pair** $\boldsymbol{A}_{i,j} = [\boldsymbol{a}_i \mid \boldsymbol{a}_j]$ spread.



STRENGTHENING THE BOUND – the RIP

Incoherence: **Each pair** $\boldsymbol{A}_{i,j} = [\boldsymbol{a}_i \mid \boldsymbol{a}_j]$ spread.



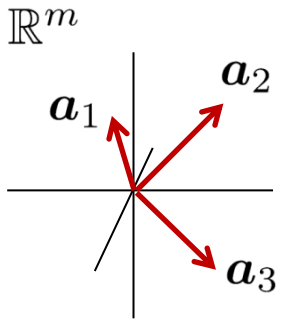
Generalize to **subsets of size** k**:**

$\boldsymbol{A}_I$ well-spread (almost orthonormal) for all I of size k

$\Longrightarrow$ all k-sparse $\boldsymbol{x}$, $\|\boldsymbol{A}\boldsymbol{x}\|_2 \approx \|\boldsymbol{x}\|_2$

STRENGTHENING THE BOUND – the RIP

Incoherence: **Each pair** $\boldsymbol{A}_{i,j} = [\boldsymbol{a}_i \mid \boldsymbol{a}_j]$ spread.



Generalize to **subsets of size** k**:**

$\boldsymbol{A}_I$ well-spread (almost orthonormal) for all I of size k

$\Longrightarrow$ all k-sparse $\boldsymbol{x}$, $\|\boldsymbol{A}\boldsymbol{x}\|_2 \approx \|\boldsymbol{x}\|_2$

$\boldsymbol{A}$ satisfies the **Restricted Isometry Property** of order k with constant δ if for all k-sparse $\boldsymbol{x}$,

$$(1-\delta)\|\boldsymbol{x}\|_2^2 \;\leq\; \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \;\leq\; (1+\delta)\|\boldsymbol{x}\|_2^2.$$

IMPLICATIONS OF RIP

Good sparse recovery

Theorem 2 (Candès+Tao '05, Candès '08) .
Suppose $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ *with*

$$\delta_{2\|\boldsymbol{x}_0\|_0} < \sqrt{2} - 1.$$

Then $\boldsymbol{x}_0$ *is the unique optimal solution to*

$$\text{minimize } \|\boldsymbol{x}\|_1 \quad \textit{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

IMPLICATIONS OF RIP

Good sparse recovery

Theorem 2 (Candès+Tao '05, Candès '08) .
Suppose $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ *with*

$$\delta_{2\|\boldsymbol{x}_0\|_0} < \sqrt{2} - 1.$$

Then $\boldsymbol{x}_0$ *is the unique optimal solution to*

$$\text{minimize } \|\boldsymbol{x}\|_1 \quad \textit{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

Again, if ... $\boldsymbol{x}_0$ **is "structured" and** $\boldsymbol{A}$ **is "nice"**

we exactly recover $\boldsymbol{x}_0$**.**

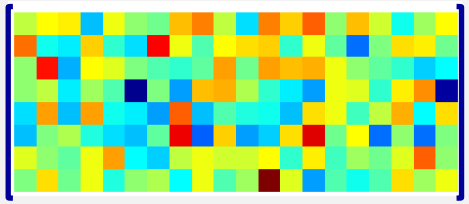
Compare condition to condition $\|\boldsymbol{x}_0\|_0 < \frac{1}{2}(1 + \mu(\boldsymbol{A})^{-1})$

IMPLICATIONS OF RIP

Random A are great:

If $\boldsymbol{A} \sim_{iid} \mathcal{N}(0, m^{-1/2})$, then

$\boldsymbol{A}$ has RIP of order k with high probability, when $m \geq Ck \log(n/k)$.

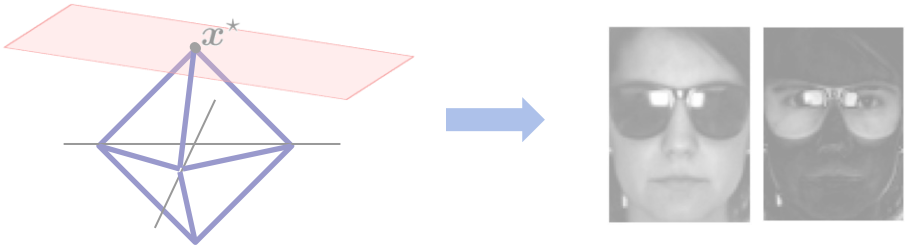


For random $\boldsymbol{A}$, ℓ^1 works even when $\|\boldsymbol{x}_0\|_0 \sim m$.

Useful property for designing sampling operators (***Compressed sensing***).

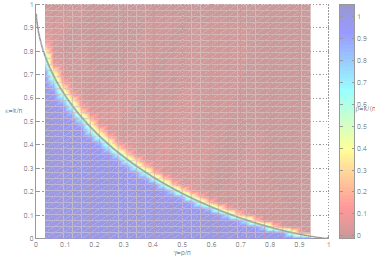
WHY CARE ABOUT THE THEORY?

Motivates applications



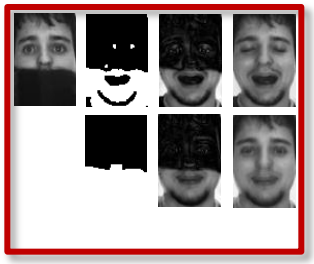
… but be careful: need to justify (and modify) the basic models [Lecture 2].

Template for stronger results



… predictions can be very sharp in high dimensions.

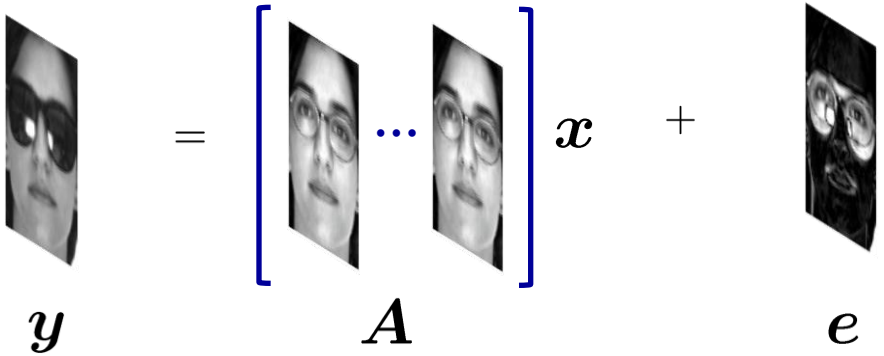
Generalizes to many other types of low-dimensional structure



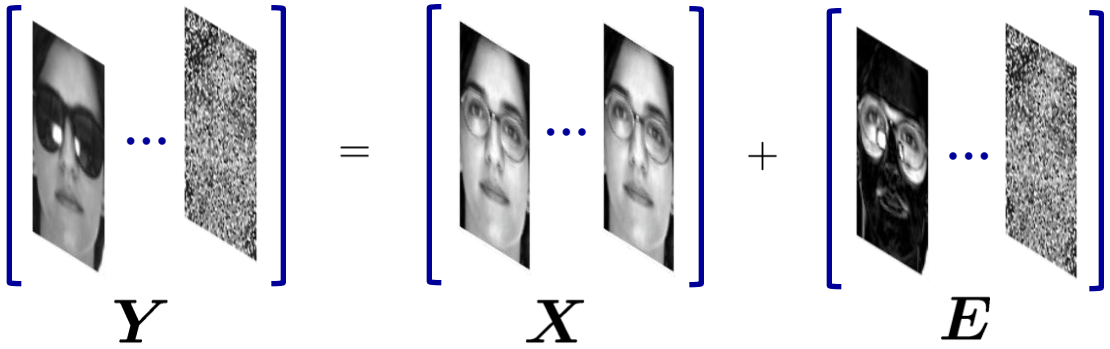
… structured sparsity, low-rank recovery [later, Lecture 2]

GENERALIZATIONS – From Sparse to Low-Rank

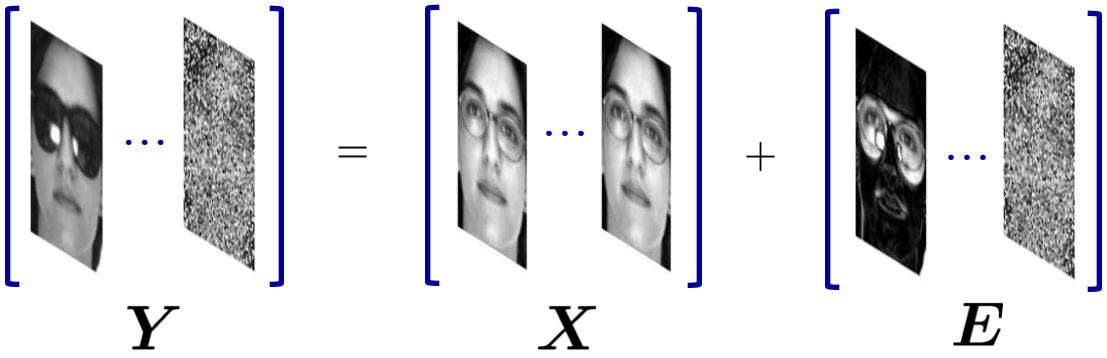
So far: Recovering a single sparse vector:



Next: Recovering low-rank matrix (many correlated vectors):



FORMULATION – Robust PCA?



Given $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{E}$, with $\boldsymbol{X}$ low-rank, $\boldsymbol{E}$ sparse, recover $\boldsymbol{X}$.

Numerous approaches to **robust PCA** in the literature:

- Multivariate trimming *[Gnanadeskian + Kettering '72]*
- Random sampling *[Fischler + Bolles '81]*
- Alternating minimization *[Ke + Kanade '03]*
- Influence functions *[de la Torre + Black '03]*

Can we give an efficient, provably correct algorithm?

RELATED SOLUTIONS – Matrix recovery

Classical PCA/SVD – low rank + noise *[Hotelling '35, Karhunen+Loeve '72,…]*

Given $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}$, recover $\boldsymbol{X}$.

Stable, efficient algorithm, theoretically optimal $\rightarrow$ huge impact

Matrix Completion – low rank, missing data

[Candès + Recht '08, Candès + Tao '09, Keshevan, Oh, Montanari '09, Gross '09, Ravikumar and Wainwright '10]

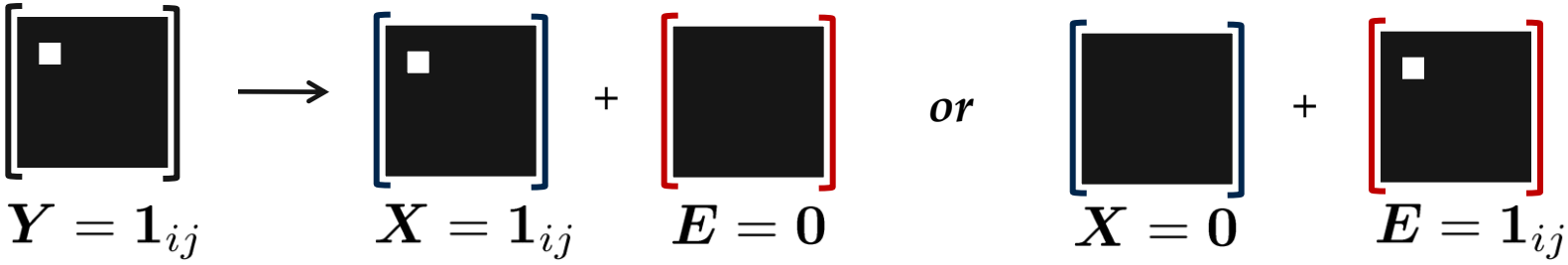
From $\boldsymbol{Y} = \mathcal{P}_{\Omega}[\boldsymbol{X}]$, recover $\boldsymbol{X}$.

Increasingly well-understood; solvable if $\boldsymbol{X}$ is low rank and Ω large enough.

Our problem, with $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{E}$, looks more difficult…

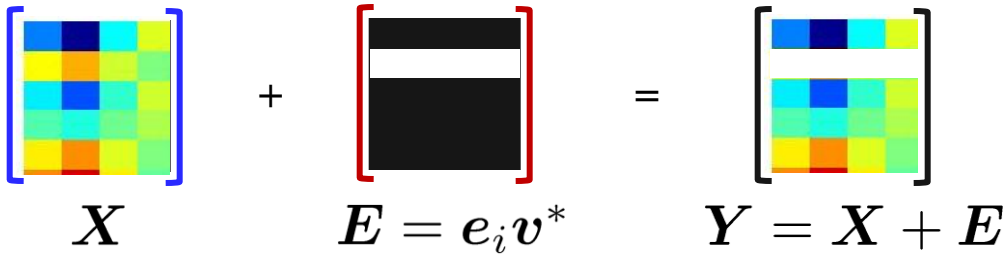
WHY IS THE PROBLEM HARD?

Some very sparse matrices are also low-rank:



Can we recover $\boldsymbol{X}$ *that are* ***incoherent*** *with the standard basis?*

Certain sparse error patterns $\boldsymbol{E}$ ***make recovering*** $\boldsymbol{X}$ ***impossible:***



Can we correct $\boldsymbol{E}$ *whose support is not* ***adversarial****?*

WHEN IS THERE HOPE? Again, (in)coherence

*Can we recover $\boldsymbol{X}$ that are **incoherent** with the standard basis from **almost all** errors $\boldsymbol{E}$?*

Incoherence condition on singular vectors, **singular values arbitrary:**

Singular vectors of $\boldsymbol{X}$ not too spiky: $\begin{cases} \max_i \|\boldsymbol{U}_i\|^2 \le \mu r/m. \\ \max_i \|\boldsymbol{V}_i\|^2 \le \mu r/n. \end{cases}$

not too cross-correlated: $\|\boldsymbol{U}\boldsymbol{V}^*\|_\infty \le \sqrt{\mu r/mn}$

Uniform model on error support, **signs and magnitudes arbitrary:**

$$\text{support}(\boldsymbol{E}) \sim \text{uni}\binom{[m]\times[n]}{\rho mn}$$

Incoherence condition: [Candès + Recht '08]

… AND HOW SHOULD WE SOLVE IT?

Naïve optimization approach

Look for a low-rank $\boldsymbol{X}$ that agrees with the data up to some sparse error $\boldsymbol{E}$:

$$\min \ \operatorname{rank}(\boldsymbol{X}) + \gamma \|\boldsymbol{E}\|_0 \quad \text{subj} \quad \boldsymbol{X} + \boldsymbol{E} = \boldsymbol{Y}.$$

$$\operatorname{rank}(\boldsymbol{X}) = \#\{\sigma_i(\boldsymbol{X}) \neq 0\}. \qquad \|\boldsymbol{E}\|_0 = \#\{\boldsymbol{E}_{ij} \neq 0\}.$$

… AND HOW SHOULD WE SOLVE IT?

Naive optimization approach

Look for a low-rank $\boldsymbol{X}$ that agrees with the data up to some sparse error $\boldsymbol{E}$:

$$\min \ \operatorname{rank}(\boldsymbol{X}) + \gamma \|\boldsymbol{E}\|_0 \quad \text{subj} \quad \boldsymbol{X} + \boldsymbol{E} = \boldsymbol{Y}.$$

$$\operatorname{rank}(\boldsymbol{X}) = \#\{\sigma_i(\boldsymbol{X}) \neq 0\}. \qquad \|\boldsymbol{E}\|_0 = \#\{\boldsymbol{E}_{ij} \neq 0\}.$$

INTRACTABLE

… AND HOW SHOULD WE SOLVE IT?

Naïve optimization approach

Look for a low-rank $\boldsymbol{X}$ that agrees with the data up to some sparse error $\boldsymbol{E}$:

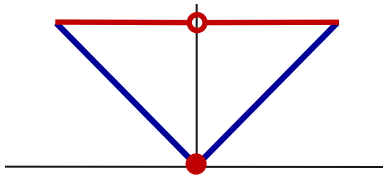
$$\min \ \operatorname{rank}(\boldsymbol{X}) + \gamma \|\boldsymbol{E}\|_0 \ \ \text{subj} \ \ \boldsymbol{X} + \boldsymbol{E} = \boldsymbol{Y}.$$

Convex relaxation

$$\operatorname{rank}(\boldsymbol{X}) = \#\{\sigma_i(\boldsymbol{X}) \neq 0\}. \qquad \|\boldsymbol{E}\|_0 = \#\{\boldsymbol{E}_{ij} \neq 0\}.$$

$$\downarrow\downarrow \qquad\qquad \downarrow\downarrow$$

$$\|\boldsymbol{X}\|_* = \textstyle\sum_i \sigma_i(\boldsymbol{X}). \qquad \|\boldsymbol{E}\|_1 = \textstyle\sum_{ij} |\boldsymbol{E}_{ij}|.$$



Nuclear norm heuristic: [Fazel, Hindi, Boyd ‘01], see also [Recht, Fazel, Parillo ‘08]

MAIN RESULT – Correct recovery

Theorem 1 (Principal Component Pursuit). *If* $\boldsymbol{X}_0 \in \mathbb{R}^{m \times n}$, $m \geq n$ *has rank*

$$r \;\leq\; \rho_r \frac{n}{\mu \log^2(m)}$$

and $\boldsymbol{E}_0$ *has Bernoulli support with error probability* $\rho \leq \rho_s^\star$, *then with very high probability*

$$(\boldsymbol{X}_0, \boldsymbol{E}_0) \;=\; \arg\min \|\boldsymbol{X}\|_* + \frac{1}{\sqrt{m}} \|\boldsymbol{E}\|_1 \quad \text{subj} \quad \boldsymbol{X} + \boldsymbol{E} = \boldsymbol{X}_0 + \boldsymbol{E}_0,$$

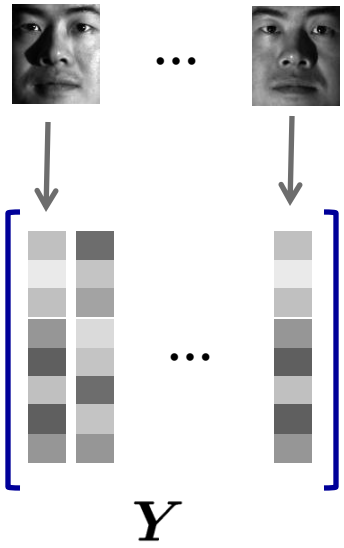
and the minimizer is unique.

"Convex optimization recovers matrices of rank $O\left(\frac{n}{\log^2 m}\right)$ ***from errors corrupting*** $O(mn)$ ***entries"***

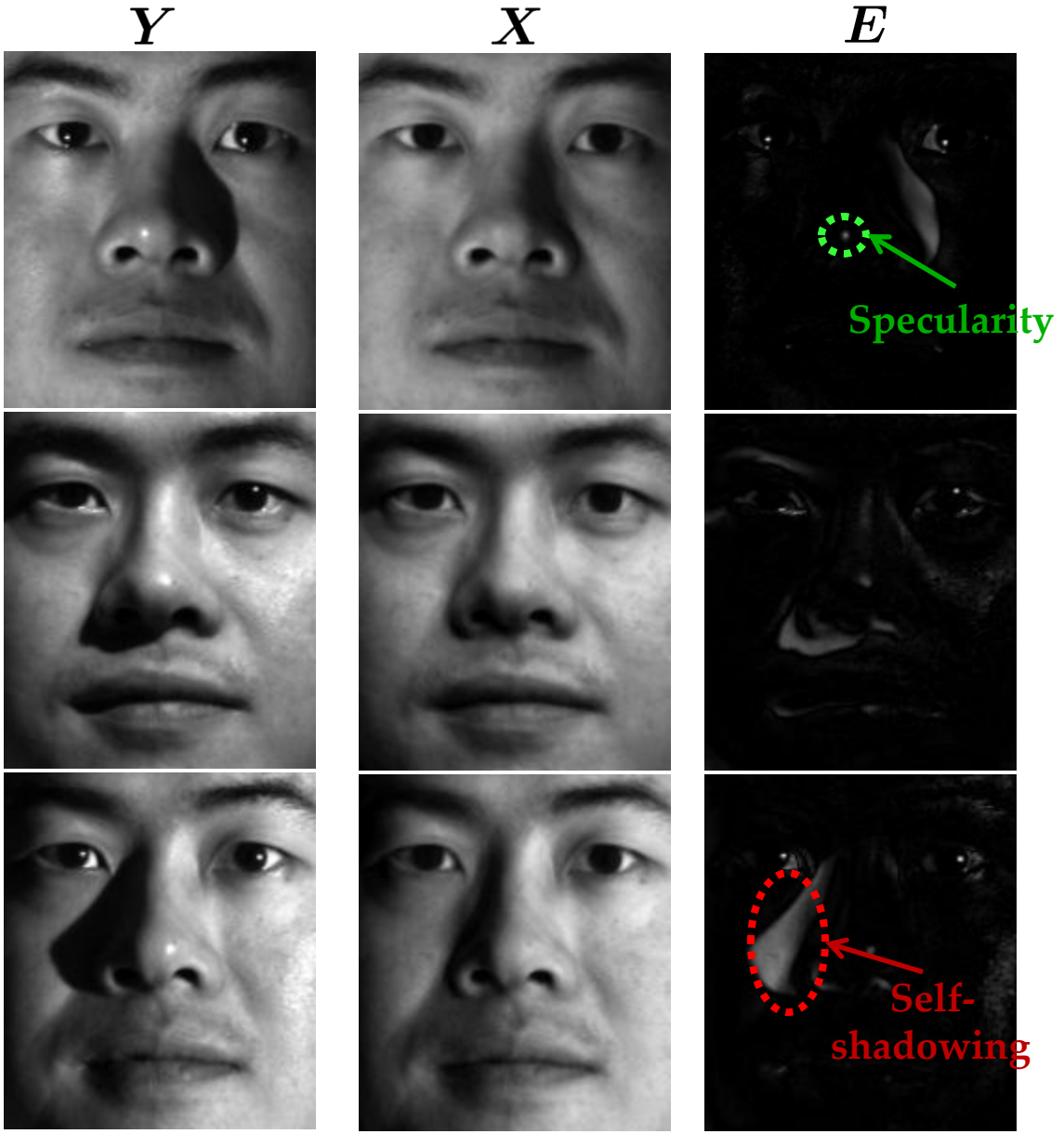
[Candès, Li, Ma, and W., '09].

EXAMPLE – Faces under varying illumination

58 images of one person under varying lighting:



RPCA

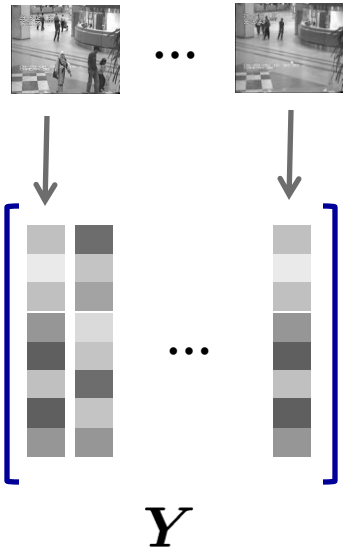


APPLICATIONS – Background modeling from video

Static camera surveillance video

200 frames, 144 x 172 pixels,

Significant foreground motion



RPCA

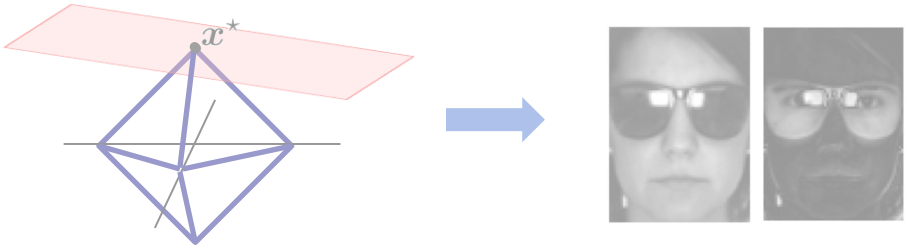
Video Y = Low-rank appx. X + Sparse error E

BIG PICTURE – Parallelism of Sparsity and Low-Rank

	Sparse Vector	*Low-Rank Matrix*
Degeneracy of	**individual signal**	**correlated signals**
Measure	**L_0 norm** $\|x\|_0$	$\mathrm{rank}(X)$
Convex Surrogate	**L_1 norm** $\|x\|_1$	**Nuclear norm** $\|X\|_*$
Compressed Sensing	$y = Ax$	$Y = A(X)$
Error Correction	$y = Ax + e$	$Y = A(X) + E$
Domain Transform	$y \circ \tau = Ax + e$	$Y \circ \tau = A(X) + E$
Mixed Structures	$Y = A(X) + B(E) + Z$	

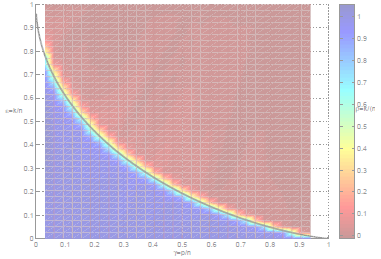
WHY CARE ABOUT THE THEORY?

Motivates applications



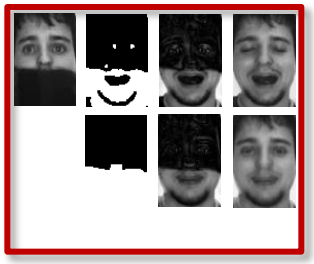
… but be careful: need to justify (and modify) the basic models [Lecture 2].

Template for stronger results



… predictions can be very sharp in high dimensions.

Generalizes to many other types of low-dimensional structure



… structured sparsity, low-rank recovery [later, Lecture 2]

General theory: constructing norms

Atomic norm: choose a set of atoms $\mathcal{A}$. Write

$$\|\boldsymbol{x}\|_{\diamond} = \inf \{\textstyle\sum_i c_i \mid \sum_i c_i \boldsymbol{a}_i = \boldsymbol{x},\ c_i > 0, \boldsymbol{a}_i \in \mathcal{A}\}$$

[Chandrasekharan et. al. '12]

General theory: constructing norms

Atomic norm: choose a set of atoms $\mathcal{A}$. Write

$$\|\boldsymbol{x}\|_\diamond = \inf \{\textstyle\sum_i c_i \mid \sum_i c_i \boldsymbol{a}_i = \boldsymbol{x},\ c_i > 0, \boldsymbol{a}_i \in \mathcal{A}\}$$

E.g., **sparsity** $\mathcal{A} = \{\boldsymbol{e}_i \mid i = 1 \dots n\},\ \|\boldsymbol{x}\|_\diamond = \|\boldsymbol{x}\|_{\ell^1}$

low-rank $\mathcal{A} = \{\boldsymbol{u}\boldsymbol{v}^* \mid \|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1\},\ \|\boldsymbol{x}\|_\diamond = \|\boldsymbol{x}\|_*$

[Chandrasekharan et. al. '12]

General theory: constructing norms

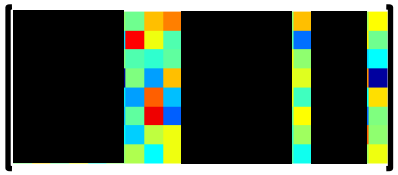
Atomic norm: choose a set of atoms $\mathcal{A}$. Write

$$\|\boldsymbol{x}\|_\diamond = \inf \{\textstyle\sum_i c_i \mid \sum_i c_i \boldsymbol{a}_i = \boldsymbol{x},\ c_i > 0, \boldsymbol{a}_i \in \mathcal{A}\}$$

E.g., **sparsity** $\mathcal{A} = \{\boldsymbol{e}_i \mid i = 1 \dots n\},\ \|\boldsymbol{x}\|_\diamond = \|\boldsymbol{x}\|_{\ell^1}$

low-rank $\mathcal{A} = \{\boldsymbol{u}\boldsymbol{v}^* \mid \|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1\},\ \|\boldsymbol{x}\|_\diamond = \|\boldsymbol{x}\|_*$

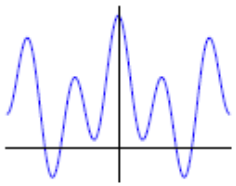
column sparsity



$$\mathcal{A} = \{\boldsymbol{u}\boldsymbol{e}_i^* \mid \|\boldsymbol{u}\|_2 = 1,\ i = 1 \dots n\}$$

e.g., [Xu+Caramanis+Sanghavi'12]

sinusoids



$$\mathcal{A} = \{e^{2\pi f t + \xi} \mid f \in [0, 1],\ \xi \in [0, 2\pi)\}$$

[Tang + Recht '12]
[Candes + Fernandez-Garza '12]

General theory: constructing norms

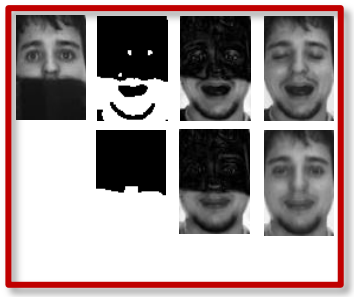
Atomic norm: choose a set of atoms $\mathcal{A}$. Write

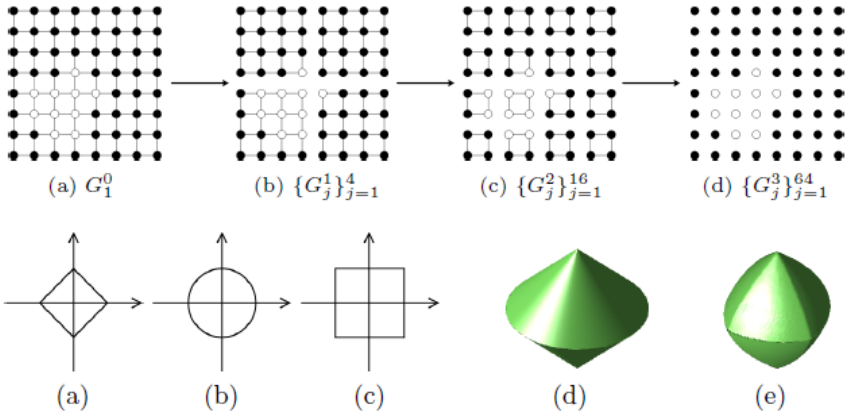
$$\|\boldsymbol{x}\|_\diamond = \inf \{\textstyle\sum_i c_i \mid \sum_i c_i \boldsymbol{a}_i = \boldsymbol{x},\ c_i > 0, \boldsymbol{a}_i \in \mathcal{A}\}$$

E.g., **sparsity** $\mathcal{A} = \{\boldsymbol{e}_i \mid i = 1 \dots n\},\ \|\boldsymbol{x}\|_\diamond = \|\boldsymbol{x}\|_{\ell^1}$

low-rank $\mathcal{A} = \{\boldsymbol{u}\boldsymbol{v}^* \mid \|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1\},\ \|\boldsymbol{x}\|_\diamond = \|\boldsymbol{x}\|_*$

spatial sparsity





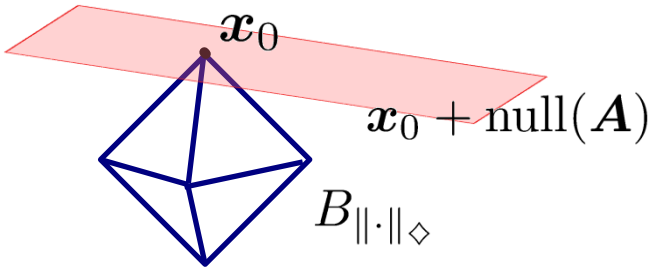
[Bach '11]
[Jia et. '12]

General theory: recovering a single structure

Observe: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ with $\boldsymbol{A} \sim \mathcal{N}(0,1)$ random. When does

$$\min \|\boldsymbol{x}\|_\diamond \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

uniquely recover $\boldsymbol{x}_0$?

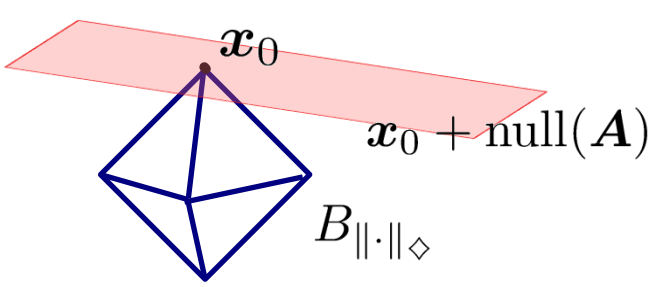


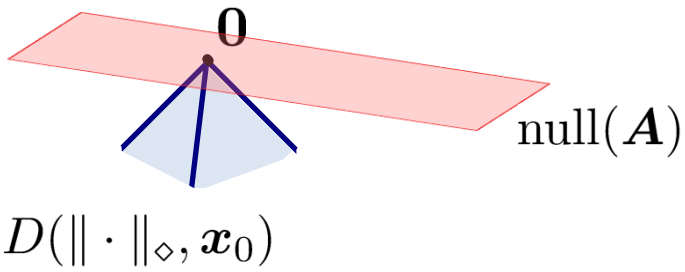
General theory: recovering a single structure

Observe: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ with $\boldsymbol{A} \sim \mathcal{N}(0,1)$ random. When does

$$\min \|\boldsymbol{x}\|_\diamond \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

uniquely recover $\boldsymbol{x}_0$?



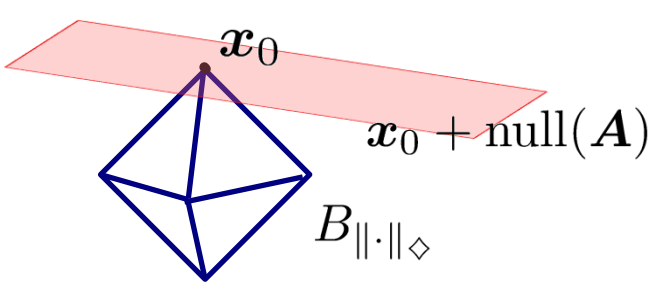


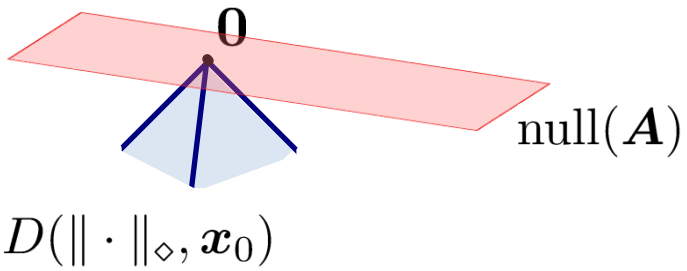
General theory: recovering a single structure

Observe: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ with $\boldsymbol{A} \sim \mathcal{N}(0,1)$ random. When does

$$\min \|\boldsymbol{x}\|_\diamond \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

uniquely recover $\boldsymbol{x}_0$?





Recovery iff the **descent cone**

$$D(\|\cdot\|_\diamond, \boldsymbol{x}_0) = \{\boldsymbol{v} \mid \|\boldsymbol{x}_0 + t\boldsymbol{v}\|_\diamond \leq \|\boldsymbol{x}_0\|_\diamond \text{ for some } t > 0\}$$

has $D(\|\cdot\|_\diamond, \boldsymbol{x}_0) \cap \text{null}(\boldsymbol{A}) = \{\mathbf{0}\}$.

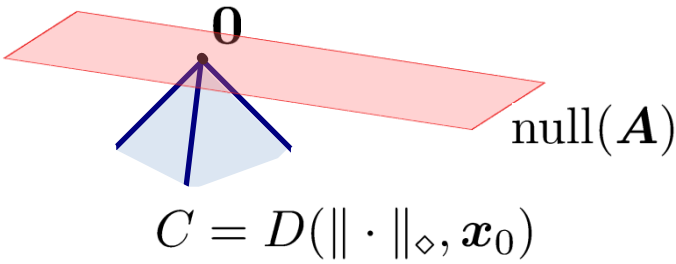
More likely if descent cone is "small". Can we make this precise?

General theory: recovering a single structure

Observe: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ with $\boldsymbol{A} \sim \mathcal{N}(0,1)$ random. When does

$$\min \|\boldsymbol{x}\|_\diamond \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

uniquely recover $\boldsymbol{x}_0$?



The **statistical dimension** of a cone C is

$$\delta(C) = \mathbb{E}_{\boldsymbol{g}\sim\mathcal{N}(0,1)}\left[\|P_C\boldsymbol{g}\|^2\right].$$

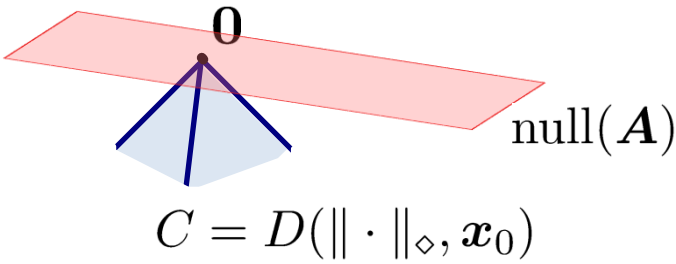
Many nice properties. E.g., if C a subspace, $\delta(C) = \dim(C)$.

General theory: recovering a single structure

Observe: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ with $\boldsymbol{A} \sim \mathcal{N}(0,1)$ random. When does

$$\min \|\boldsymbol{x}\|_\diamond \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

uniquely recover $\boldsymbol{x}_0$?



The **statistical dimension** of a cone C is

$$\delta(C) = \mathbb{E}_{\boldsymbol{g}\sim\mathcal{N}(0,1)}\left[\|P_C\boldsymbol{g}\|^2\right].$$

Many nice properties. E.g., if C a subspace, $\delta(C) = \dim(C)$.

Sharp **phase transition** at $m = \delta(C)$:

$$\begin{aligned} m > \delta(C) \quad &\Longrightarrow \quad \mathbb{P}[\text{recovery}] > 1 - \exp\left(-c(m-\delta(C))^2/n\right) \\ m < \delta(C) \quad &\Longrightarrow \quad \mathbb{P}[\text{recovery}] < \exp\left(-c(m-\delta(C))^2/n\right) \end{aligned}$$

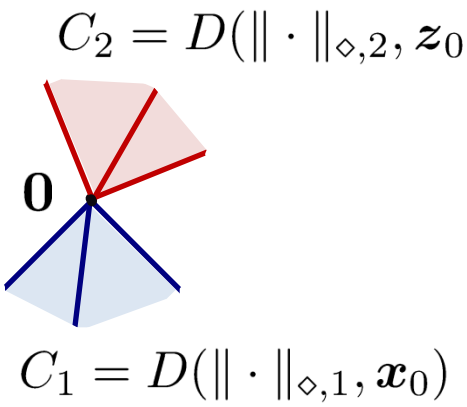
[Amelunxen, McCoy, Lotz, Tropp '13]

General theory: decomposing two structures

Observe: $\boldsymbol{y} = \boldsymbol{x}_0 + \boldsymbol{z}_0$ with regularizers $\|\boldsymbol{x}\|_{\diamond,1}$, $\|\boldsymbol{z}\|_{\diamond,2}$. Does

$$\min \|\boldsymbol{x}\|_{\diamond,1} + \|\boldsymbol{z}\|_{\diamond,2} \quad \text{s.t.} \quad \boldsymbol{x} + \boldsymbol{z} = \boldsymbol{y}$$

uniquely recover $\boldsymbol{x}_0, \boldsymbol{z}_0$?



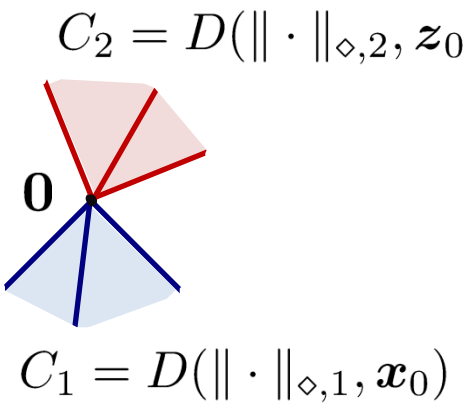
Variant: $\min \|\boldsymbol{x}\|_{\diamond,1} \quad \text{s.t.} \quad \|\boldsymbol{z}\|_{\diamond,2} \leq 1, \ \boldsymbol{x} + \boldsymbol{z} = \boldsymbol{y}$

General theory: decomposing two structures

Observe: $\boldsymbol{y} = \boldsymbol{x}_0 + \boldsymbol{z}_0$ with regularizers $\|\boldsymbol{x}\|_{\diamond,1}$, $\|\boldsymbol{z}\|_{\diamond,2}$. Does

$$\min \|\boldsymbol{x}\|_{\diamond,1} + \|\boldsymbol{z}\|_{\diamond,2} \quad \text{s.t.} \quad \boldsymbol{x} + \boldsymbol{z} = \boldsymbol{y}$$

uniquely recover $\boldsymbol{x}_0, \boldsymbol{z}_0$?



Variant: $\min \|\boldsymbol{x}\|_{\diamond,1} \quad \text{s.t.} \quad \|\boldsymbol{z}\|_{\diamond,2} \leq 1, \ \boldsymbol{x} + \boldsymbol{z} = \boldsymbol{y}$

In a random incoherence model (C_2 randomly rotated), **phase transition** at

$$\delta(C_1) + \delta(C_2) = n$$

$$\begin{aligned} n > \delta(C_1) + \delta(C_2) &\implies \mathbb{P}[\text{recovery}] > 1 - \exp\left(-c(n - \delta(C_1) - \delta(C_2))^2/n\right) \\ n < \delta(C_1) + \delta(C_2) &\implies \mathbb{P}[\text{recovery}] < \exp\left(-c(n - \delta(C_1) - \delta(C_2))^2/n\right) \end{aligned}$$

[Amelunxen, McCoy, Lotz, Tropp '13]

General theory: statistical estimation

Observe: noisy measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{z}$. Noise-aware optimization:

$$\min \|\boldsymbol{x}\|_\diamond + \tfrac{\gamma}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$$

E.g., **Basis pursuit denoising:** $\min \ \|\boldsymbol{x}\|_1 \ + \ \tfrac{\lambda}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$

Noise-aware RPCA: $\min \ \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{S}\|_1 + \tfrac{\gamma}{2}\|\boldsymbol{L} + \boldsymbol{S} - \boldsymbol{D}\|_F^2$

When does $\min \|\boldsymbol{x}\|_\diamond + \tfrac{\gamma}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$ produce $\hat{\boldsymbol{x}} \approx \boldsymbol{x}_0$?

General theory for **decomposable regularizers** $\|\cdot\|_\diamond$

[Negahbhan, Agarwal, Yu, Wainwright '12]

A suite of models and theoretical guarantees

For robust recovery of a family of low-dimensional structures:

- [Zhou et. al. '09] **Spatially contiguous** sparse errors via MRF
- [Bach '10] – structured relaxations from **submodular functions**
- [Negahban+Yu+Wainwright '10] – **geometric analysis** of recovery
- [Becker+Candès+Grant '10] – **algorithmic templates**
- [Xu+Caramanis+Sanghavi '11] **column sparse errors $L_{2,1}$ norm**
- [Recht+Parillo+Chandrasekaran+Wilsky '11] – **compressive sensing of various structures**
- **[Candes+Recht '11] –** **compressive sensing of decomposable structures**

$$X^0 = \arg\min \|X\|_\diamond \quad \text{s.t.} \quad \mathcal{P}_Q(X) = \mathcal{P}_Q(X^0)$$

- **[McCoy+Tropp'11] –** **decomposition of sparse and low-rank structures**

$$(X_1^0, X_2^0) = \arg\min \|X_1\|_{(1)} + \lambda \|X_2\|_{(2)} \ \text{ s.t. } \ X_1 + X_2 = X_1^0 + X_2^0$$

- **[W.+Ganesh+Min+Ma, I&I'13] –** **superposition of decomposable structures**

$$(X_1^0, \ldots, X_k^0) = \arg\min \sum \lambda_i \|X_i\|_{(i)} \ \text{ s.t. } \ \mathcal{P}_Q(\textstyle\sum_i X_i) = \mathcal{P}_Q(\sum_i X_i^0)$$

Take home message: ***Let the data and application tell you the structure…***

Thank you! Question time …

Next: **Modeling** low-dimensional structure in **real data.**

Later: Solving the **optimization** problems efficiently.

Tomorrow: A **Bayesian** way of approaching sparse modeling.